Numerical Competence in a Response Reproduction Task

A thesis
Submitted in partial fulfillment
of the requirements for the degree
of
Master of Arts in Psychology
at the
University of Canterbury
by Shasta Holland

University of Canterbury
March 2001

QL
785.24
H737
2001

ACKNOWLEDGMENTS

Most of all I would like to thank my supervisors Anthony McLean and Randy Grace for their support and suggestions. They have both been incredibly helpful. I realize that my being in Australia for the final month has made their roles much more difficult, and I really appreciate the extra effort they have gone to on my behalf.

Also, a big thanks to Marion and Fiona for their work in the lab, and for our Friday nights out. I hope they will continue.

Finally, to my parents I don't know where I would be without you.

20 NOV 2001

TABLE OF CONTENTS

FIGURES

TABLES

ABSTRACT

A response reproduction paradigm for investigating numerical ability in animals is presented. In this experiment subjects were presented a number of light flashes (2, 4, or 6) in a sample phase, and were then required to reproduce that number in a production phase by pecking one key either 2, 4 or 6 times and then pecking another key once. Flashes in the sample phase were presented in one of two manners. Either one flash occurred every 2.5-seconds (termed the rate-controlled procedure), or all flashes occurred within a 10-second interval (termed the time-controlled procedure). Two subjects were trained in each procedure. Following training, testing with novel probe trials (1, 3, 5 and 7-flash sequences) was conducted. Novel probe trials, interspersed with baseline trials, were either presented in the same manner as the baseline trials, consistent transfer, or in the opposite procedure to baseline trials, inconsistent transfer. Reinforcement was delivered in a random manner on probe trials, and only for correct responses on baseline trials. In a second condition training and testing conditions were reversed, so that subjects that had received time-controlled training now received rate-controlled training and vice versa. All subjects learned to respond in the task with reasonable accuracy. Response distributions for the three baseline trials were distinct from one another and peaked at, or near, the reinforced number of responses. The introduction of novel probe trials disrupted baseline performance to varying degrees. In consistent transfer testing number of responses increased as flash number increased in an orderly manner for all subjects. During inconsistent transfer testing differences between rate and time-controlled performance emerged. All subjects maintained to some degree their baseline trial performance, however, time controlled subjects showed little discrimination between probe trials, while rate-controlled subjects made fewer responses to trials consisting of large numbers of flashes (5 and 7) than to trials consisting of small numbers of flashes (1 and 3). Examination of standard deviations of responses and coefficients of variation suggest that subjects in the rate-controlled procedure were relying upon temporal cues. Subjects Q5 and Q8 in the time-controlled procedure also appear to have used temporal cues. A confusion-diffusion model describing subjects' performance in this experiment is also presented. The usefulness of the response reproduction paradigm for investigating animal numerical ability is discussed and modifications to the present procedure are suggested.

INTRODUCTION

Can animals count? Along with other cognitive processes that have been assumed to lie exclusively in the human domain, the question of whether animals are capable of numerical competence has intrigued researchers for more than a century. Interest in the field of animal cognition in general, and animal numerical competence in particular, has been steadily increasing in the last few decades. Research in this area seems to have arisen from two distinct backgrounds. First are those researchers who are primarily interested in understanding the development of human numerical ability and believe that animal processes may provide some insight. Other researchers are more interested in animal numerical competence in itself, without reference to its implications for understanding human abilities.

Counting in Humans

The most frequently cited definition of counting in the animal cognition literature comes from Gelman and Gallistel (1978). They formulated their definition with the view that a comprehensive definition of counting should be capable of including non- or pre-verbal humans, such as very young children. Gelman and Gallistel state that it may be inappropriate to use adult standards as a measure of a child's ability to count. Although adults use number words in a specific order, the inability of a child to do the same does not necessarily preclude them from "counting".

Support for this view comes from studies of counting in different cultures. Gelman and Gallistel state that the belief that many African tribes cannot count could have arisen from researchers failing to recognize that hand gestures, which were either unrelated or inconsistently related to words, represent numbers. Similarly, if verbal labels are used they may bear no relation to European number words. Gelman and Gallistel provide an example involving South Sea Islanders who use a base two number system. It may appear that the numbers one and two are the only numbers available to these people, and yet they have the ability to represent numerosities greater than three. These examples illustrate the problem of requiring a standard verbal representation of number for counting to be said to occur. With such a definition many forms of counting may not be recognised, resulting in the erroneous conclusion that those individuals are unable to count. Thus, any definition of counting must be able to include counting that uses a system of enumeration that is not necessarily based on a verbal standard.

Gelman and Gallistel (1978) developed five principles that they state are necessary for counting to occur. The first four principles involve the procedure of counting, or how to count. Firstly, each item to be counted must be ticked off or associated with a distinct numerical tag. This principle is known as the *one-to-one correspondence principle*. The application of this principle involves two processes, *partitioning* and *tagging*. Partitioning is the separation of items that have been counted from those that are yet to be counted, and involves the movement of items from one category to the other as the counting process continues. Tagging is the application of a distinct tag to each of the items that are being counted. Verbal tagging, however, is not necessary to fulfil this condition; any series of symbols or

behaviour is adequate. The process by which items are tagged may even be the activation of nodes in short-term memory (Davis and Memmott, 1982). Kohler (1950, in Davis and Perusse, 1988) proposed that animals might use a system of inner marks to 'think unnamed numbers'.

Once an item has had a tag applied to it, the item is moved from the yet-to-be-counted category to the set of items that have been counted, and that tag is then unavailable for subsequent use with that set. In order for a correct count to be achieved, partitioning and tagging must start at the same time and end together.

The second principle is called the *stable-order principle* and requires that the tags available for use be applied in a fixed order. Gelman and Gallistel note that it is this principle that provides the greatest difficulty for children. Because verbal tags are arbitrary in nature, the ability of a child to count a set of items relies upon their ability to remember a given sequence of these verbal tags, a task that increases in difficulty as the set size increases.

The third principle is the *Cardinal principle*, which states that the final tag in a series represents the numerosity of the whole set. Besides as being able to apply tags to a set of items, the individual must recognise this descriptive property of the cardinal tag. It should be noted that cardinality is necessary but not sufficient for counting. It is possible to learn a set of tags without knowledge of the ordered relationship between them.

The *order-irrelevance principle* states that it makes no difference which item is tagged first, as long as each item is tagged only once. Understanding this principle involves an implicit understanding that the cardinal number of a set remains the same for any order of enumeration, that each tag is temporarily applied to an item, and that the tags are independent of the items themselves.

The final principle gives the rule governing what to count. The *abstraction principle* states that the preceding four principles can be applied to any set of entities, whether physical or non-physical. Together, these principles form the requisites for counting to occur.

The Spectrum of Numerical Competence

Davis and Perusse (1988) sought to clarify research on numerical competence in non-humans by providing concise definitions of the range of numerical abilities that are discussed in the animal cognition literature. It is commonly assumed that these abilities lie on a continuum ranging from simple judgements of relative numerosity to the higher cognitive processes of counting and a concept of number. *Relative numerousness judgements* are dichotomous, more versus less, comparisons that do not involve knowledge of absolute number. It has been suggested that this ability may be the starting point for more formal enumerative processes such as counting.

The next numerical ability on the continuum is termed *subsitising* and involves the rapid assignment of numerical tags to a small range of numbers, typically one to six. This process is generally thought to be perceptual rather than cognitive or enumerative in nature, and is based on pattern recognition rather than a formal knowledge of numbers. Davis and Perusse (1988) note that children can learn the first few number words as names of object configurations before they have any numerical concepts. For example, a child may know that the word three describes a picture of three cups, without knowing the relation of the word three to the number '3'.

There is some disagreement regarding this definition however. Several researchers have proposed that, in humans at least, subsitising may develop after full counting has been achieved, rather than before (e.g., Gelman and Gallistel, 1978; Gallistel, 1988; Miller, 1993). When human adults are asked to state as quickly as possible how many items are in a display, reaction times for numbers one to four are very fast and accurate. When the number of items in the display increases above four, reaction times also increase in a systematic fashion, as do the number of errors. This suggests that the processes used to determine the number of items differ over the two ranges of numerosities. Subsitising is said to be responsible for discriminations with small numbers and estimation for those numbers greater than four or five. If subsitising is a precursor to full numerical competence, then children who are still learning to count should show rapid and accurate discrimination of small numbers just as adults do. However, Gelman and Gallistel (1978) found no evidence of this and instead proposed that subsitising arises out of counting competency. According to their view, once we have had considerable practice at counting small numbers we begin to recognise the patterns that these numbers form, and hence discrimination transfers from the domain of cognition to perception. Thus, although subsitising appears to be a post-counting ability in humans, we cannot assume that this must also be the case for non-humans. It may be that for non-humans there is a fast perceptual ability that precedes counting. However, there does not appear to be much empirical evidence that supports this idea.

Researchers in numerical competence have adopted Gelman and Gallistel's (1978) definition of counting for use with non-humans. In fact Gelman and Gallistel themselves proposed that their set of principles were applicable for non-humans. They state that there is nothing in their definition that precludes non-verbal humans, and by

extension non-verbal non-humans, from counting. The applicability of this extension is an issue that will be discussed later. One point worth noting, however, is that Gelman and Gallistel's abstraction principle bears remarkable similarity to the transfer of training requirement often used in animal experimentation. That is, in both cases, the individual must be able to take the principles/processes used in one situation and apply them in another novel situation. If this requirement is not met then true learning of these principles/processes is said not to have occurred.

Another category of numerical ability proposed by Davis and Perusse is that of *protocounting*. The protocounting category encompasses research in which counting is the most likely numerical process but not all control tests have been carried out, such as when transfer of ability across situations has not been demonstrated. Davis and Perusse state that at present all counting research with animals falls into the protocounting category, with the possible exception of some primate literature. While many experiments are suggestive of counting they do not fulfill all of Gelman and Gallistel's principles, most notably the ability to count arbitrary items in any situation. It should be noted that children cannot necessarily meet these criteria either, so to require this of animals may be unreasonable. However, the inclusion of the protocounting category is somewhat contentious. Several authors (Boysen, 1988; Braaten, 1988; Johnson, 1988) have claimed that an individual either can or cannot count, and if the experimenter has not demonstrated all of the necessary features little can be said regarding that individuals ability. Thus, the term protocounting provides little clarification.

Gelman and Gallistel also state that partial competence or limited ability is not the same thing as a complete lack of competence. It is not true that to be able to count one must always count correctly. In fact, Gelman and Gallistel propose that there is

variability inherent in a pre-verbal counting system, which is gradually overcome by practice with verbal labels. Thus, one would not expect non-verbal individuals to always arrive at a correct 'count' because of this variability.

Finally, Davis and Perusse have differentiated, as have other researchers, between an ability to count and *a concept of number*. The latter refers to the ability to mentally manipulate numbers without reference to objects that instantiate those numbers. This involves an understanding of the abstract and representational properties of number. Examples of these manipulations include operations such as addition and subtraction. While counting is a prerequisite for a concept of number, a concept of number is not required for counting to occur, and in humans it appears to develop after counting ability. Together these five terms encompass the various numerical abilities that can occur in different species. What remains to be determined is what degree of numerical competence these different species are capable of attaining.

Issues in the Animal Numerical Competence Area

Appropriateness of human models of counting:

A major assumption made by researchers in the animal cognition is that there is a continuum between the abilities of humans and other animals. This idea can be traced back to Darwin's (1871/1920) theory of the evolution of intelligence, where he posited a mental continuum between humans and animals:

> " The difference in mind between man and the higher animals, great as it is, certainly is one of degree and not one of kind." (p.128)

However, there is some disagreement regarding this idea with respect to numerical competence. Karmiloff-Smith (1988) states that, as with animal communication, there are fundamental differences between animal and human numerical abilities, namely they way in which number is represented. This suggests that differences in ability are qualitative and not simply quantitative, and that using a definition of enumeration based on human counting to investigate the extent of non-human numerical ability may be inappropriate. Several researchers (e.g. King, 1988; Luchins and Luchins, 1988; Davis, 1993) have suggested that by focusing on counting in a human sense, important aspects of animal numerical ability may be overlooked. The question of what animals can do and how they do it is different from the question of whether animals can count in an equivalent manner to humans. These two questions may produce different results, or alternatively suggest that there is in fact a continuum between human and non-human numerical ability.

Although some researchers believe that the underlying processes numerical ability is based on differ for humans and non-humans, others believe that a common mechanism is shared. Gallistel and Gelman (1992) have presented an interesting account of human and non-human numerical ability that rests on three basic ideas. First, a pre-verbal counting mechanism, such as a counter accumulating neural pulses proposed by Meck and Church (1983), guides the acquisition of verbal counting in humans. Second, learning to count involves, in part, the mapping of pre-verbal magnitudes to verbal and written number symbols, and vice versa. And third, the foundations of human numerical ability lie in animal pre-verbal counting and arithmetic reasoning. Gallistel and Gelman differentiate between numerons, which are mental representations of numerosity, and numerlogs, which are verbal labels for numerons. They argue that the acquisition of verbal counting involves the mapping of

numerlogs onto the pre-existing numerons. The existence of these pre-verbal numerons makes verbal counting intelligible, and hence learnable.

One piece of evidence supportive of their view comes from the difficulty that children have in learning fractions. The pre-verbal system is restricted to discrete numerons and therefore cannot represent fractional numerosities. Because of this children are unable to map fractions onto the pre-verbal system and thus have difficulty understanding them. For example, when asked to show what three thirds is equivalent to, children around six or seven years will point to the number three. Similarly, children cannot order fractions correctly, often choosing ¼ as larger than ½. They have great difficulty learning that between any two discrete numbers lies a large range of fractional numbers. This supports the idea that the learning of numerlogs and associated ideas is facilitated by the presence of pre-existing numerons. When there is no correspondence between these pre-verbal numerons and subsequent numerlogs, such as with fractions, learning is much more difficult.

Gallistel and Gelman (1992) propose that although human numerical ability is significantly different than that of non-human species, the basic mechanisms underlying those abilities are the same. They suggest that the way in which pre-verbal humans and non-human species process number is the same. However, the presence of linguistic ability in humans results in a more accurate numerical system, albeit one that is inextricably linked with the pre-verbal system. If one accepts this view, Gelman and Gallistel's (1978) definition of counting can be equally applied to humans and non-humans, because, in theory, the pre-verbal counting mechanism present in each adheres to these principles.

The naturalness of number:

A second debate in the animal numerical competence literature centers on whether the use of number is natural for animals. It could be argued that the ability of an animal to make numerical discriminations is evidence of such naturalness, however the naturalness of a skill seems to require something over and above its presence in an organism. This requirement seems to be that the organism actually uses the ability as a matter of course in its environment. There are two discrepant viewpoints with respect to this idea. Davis is the major proponent of the last-resort hypothesis (Davis and Memmott, 1982; Davis and Perusse, 1988). The last-resort hypothesis states that while non-human animals may be trained to make discriminations on the basis of number, this ability is relatively unnatural and is only used when other reliable cues are unavailable. Another consideration regarding the naturalness of number involves the ease with which these skills are acquired. Davis and Memmott (1982) argue that while the development of human numerical ability requires hundreds and thousands of trials, it is the nature of this development rather than its duration that is important. Even though numerical ability in humans develops over the course of years it does so with remarkable ease, even in cases where little effort is taken to encourage it. Thus, human counting behaviour seems to evolve inevitably, while other animals counting behaviour requires extensive input and maximal environmental support. However, the training trials that humans receive are extremely varied in content, difficulty and context, while those used in experiments with animals are typically repetitive. Variation is exactly what would be expected to give rise to abstraction and generalized imitation in animal concept formation. Transfer of "concept formation" to novel stimuli has been found to be greater with a larger number of exemplars (e.g. Wright,

1997). Thus, providing varied training with animal subjects may reduce differences in the nature of the development of numerical ability between animals and humans.

In contrast to Davis, Capaldi and Miller (1988) propose that not only do animals make numerical discriminations on the basis of number but that this ability is quite natural and routinely used in the presence of other equally valid cues. They suggest that unlike light flashes, tones or geometric shapes, reinforcing events such as food deliveries are relevant to the animal in its natural environment. Therefore, animals are likely to discriminate between reinforcements on the basis on number, when they may be unlikely to do so for other 'irrelevant' stimuli. Other researchers such as Meck and Church (1983) and Gelman and Gallistel (1978) also adhere to the view that animals naturally process numerical information. To investigate the issue of the naturalness of number discrimination several researchers have examined whether time or number is the most salient cue regulating behaviour. This research will be reviewed later.

The 'naturalness' debate also raises the question of whether it really matters that numerical information be used in preference to other forms of information. However, whether or not a particular species actually attends to numerical cues as a matter of course does not make the finding that they *are capable* of utilizing numerical cues any less interesting. This is especially true when examining the question of whether humans and other animals process numerical information in a similar fashion.

Research into Numerical Competence in Non-Human Species

There are two main types of procedures used to study numerical competence in animals. These involve either the sequential or simultaneous presentation of stimuli. In simultaneous procedures the subject is presented with two arrays of elements at the same time and is then required to discriminate between them. Generally subjects are reinforced for responses to one alternative if one array is larger and to another alternative if the other array is larger. In sequential procedures, stimuli are presented or encountered one at a time. Subjects then indicate if the proportion of one item is large or small by responding to one of two alternatives. Note that to some extent this resembles a memory task, as the stimuli are no longer present when the choice response is made.

Relative Numerosity Discriminations:

Relative numerical judgements are the simplest form of numerical competence. They involve dichotomous, more-versus-less judgements, which are not based on knowledge of absolute number. Honig and Stewart (1989) argued that the ability to make relative numerosity judgements is of particular importance to animals. Being able to discriminate the relative numbers of edible items in two different foraging areas or the number of predators is clearly advantageous to an animal's survival. In contrast, precise counting is likely to be an unnecessary and inefficient skill. It thus seems reasonable from and evolutionary perspective that animals are capable of making discriminations based on the relative numerosity of sets of items. The claim that animals can make these relative numerosity judgements implies that

they discriminate of the basis of the number of items in an array and not some other feature such as colour, shape or density.

Several researchers have examined relative numerosity in pigeons. Honig and Stewart (1989) used simultaneous arrays varying in terms of colour, form or size. In their first experiment, uniform arrays of red and blue dots (36 total) were projected onto a response panel, and differed in the proportion of red and blue dots present. Pigeons were trained with arrays of all red or all blue dots until they could discriminate the positive stimulus elements (S+) in 95% of trials. The test stimuli were arrays with varying proportions of the two colours. They found that the proportion of responses to the negative stimulus elements (S-) decreased as the number of positive elements increased, indicating that the subjects could discriminate between the different proportions of the two elements. Transfer tests involving 64 dot arrays were conducted, in which discrimination remained stable. The transfer tests indicated that relative and not absolute number guided judgements.

The authors noted that when objects are a constant size the area of the array is correlated with the number of elements in it. Consequently, the discrimination may have been based on area or size rather than numerosity. Experiment 2 was conducted to investigate this possibility. Instead of different coloured dots, crosses and zeros (X and 0) were used. Because of the disparate nature of these forms, summing over their respective areas was considered unlikely. Discrimination during both test and transfer trials was very accurate, suggesting that judgements were based on relative numerosity. Another way to test whether discriminations are based on number or area is to vary the sizes of the elements in the array. Large dots take up more area than the same number of small dots. If animals do sum area then a greater number of small dots should be perceived as the same as a small number of large dots. Colour and

shape were kept constant preventing the pigeons from using these factors in their discriminations. Discrimination was as accurate in these tests as in previous experiments, again indicating that the area of the items in the stimulus array did not control judgements.

The final experiment conducted by Honig and Stewart involved items from 'natural' categories. Various non-identical pictures of birds and flowers were chosen. Discrimination was very accurate, again suggesting that relative number and not other parameters of the stimulus array controlled discrimination judgements. These experiments, together with similar results from Emmerton (1998), indicate that pigeons can discriminate the relative numerosities of items in simultaneously presented stimulus arrays. Smirnova, Lazareva and Zorina (2000), using a matching to sample paradigm with crows have also demonstrated relative numerousness judgements.

Researchers have also presented stimuli sequentially in order to assess relative numerosity discriminations. Alsop and Honig (1991) used this procedure to examine relative numerosity in pigeons. They noted that pigeons allocate behaviour between two concurrent alternatives as a function of relative reinforcement for those alternatives (Herrnstein, 1961). This suggests that in addition to simultaneously presented events pigeons are also sensitive to the relative numbers of events that occur serially. Pigeons were required to discriminate between the relative numbers of red and blue flashes in a sequence. The total number of flashes in a sequence was held constant while the relative numbers of the red and blue flashes that comprised the sequence were varied. Once the sequence had been presented the pigeons responded on one of two keys to indicate whether the majority of flashes were red or blue. As

the number of blue flashes in the sequence increased so did the proportion of 'more blue' responses. This ability was transferred to sequences with fewer or greater numbers of elements, replicating Honig and Stewart's findings with simultaneous arrays. Another interesting finding was that responding was biased towards the colour of the flashes presented last (a recency effect). This suggested that stimuli later in the sequence exerted more control over discrimination judgements than did stimuli earlier in the sequence.

Alsop and Honig's Experiment 2 explored this 'recency' effect. The stimulus array consisted of five flashes of light where the combinations of red and blue were mixed rather than grouped as in Experiment 1. For example, BRRRR, RRBRR, BBBRB etc. Red and blue trails were combined to give a majority-odd description of the findings, e.g. MMMOM. They found that accuracy was the greatest for all red or all blue sequences. Accuracy was decreased slightly by the introduction of one odd flash at the beginning of the sequence. As the odd flash moved towards the end of the sequence accuracy consistently decreased. This is the same recency effect that was seen in Experiment 1, which suggests that position of the elements in an array is important in discrimination judgements.

However, Alsop and Honig suggested that element position may not be the controlling factor. They proposed that the time elapsed between the presentation of an element and the opportunity to respond is the critical variable. In Experiment 3 the duration of gaps between flashes were varied in three-flash sequences. If temporal location of the flash is important, then as the gap between the first and second flash is increased the influence of the first flash on choice should decrease. The results showed that for OMM trials, increasing the duration of the gap between flashes increased subject's accuracy. Conversely, in MMO trials accuracy decreased when

gap duration increased. This was what was predicted if the first flash has the least influence on discrimination. When the odd flash is first, the two majority flashes have greater influence, and are thus easily discriminated as M, conversely, when the odd flash is last it will be the most salient and will consequently be more likely to be classified as M. These results suggest that the influence of any one element rapidly decays and that memory is an important factor in an animal's performance in sequential discrimination tasks. An alternative explanation for these results is that the pigeons may not be discriminating in terms of number. Instead they may be summing the duration of each element in the sequence and responding according to which duration was longest. This possibility will be addressed later.

Overall, these experiments indicate that pigeons are capable of making discriminations on the basis of the relative number of items in an array and not other features of the stimuli such as size, shape or density.

Absolute Number Discriminations:

Absolute number discriminations, while still short of counting, are considered to be a higher order ability than relative numerousness judgements. Several researchers have examined the extent of animal's abilities to make absolute and not relative numerosity judgments. This work has been done with a variety of species including rats, pigeons and monkeys. In order to demonstrate that an animal can make an absolute number discrimination it must be shown that 1) the animal is not making a relative numerousness judgement; 2) that discrimination is based on the number of the stimuli and not area or duration; and 3) that discrimination is not stimulus specific.

Davis and Bradford (1991) have examined absolute number discrimination in rats. They noted that under natural conditions animals will eat until they are satiated or are forced to leave. Koehler (1950, cited in Davis and Bradford) trained budgies to eat a fixed number of grains, between two and six, from arrays that were larger than the target number. Davis and Bradford set out to replicate this finding using rats. Subjects were assigned a target number, ranging from 3-5, designating the number of pellets they were allowed to eat. They were trained to move onto a plank from the start area, eat the required number of pellets, and then return to the start area. Correct trials were reinforced with verbal praise, petting and a food pellet, while incorrect responses (eating more than the required number) were punished by a loud noise (verbal "NO" and a hand-clap). If subjects consumed fewer pellets than the target number neither reinforcement nor punishment were delivered. The number of pellets presented in the array was increased to between 15 and 20, and arrays were randomly varied in size and pattern across sessions.

All subjects ate the required number of pellets more often than they ate a fewer or larger number of pellets. Although considerable effort was made to control for pattern over trials, subjects' food selection did not conform to a set pattern, even when food was symmetrically placed. In Phase 2, the pellets were replaced with larger irregularly sized sunflower seeds. Performance was initially disrupted but quickly returned to previous levels. These results demonstrate that rats, who are naturally inclined to eat all available food, can be taught to restrict each meal to a target number. The transition from pellets to irregularly shaped seeds indicates that the size or volume of food items were not factors in discrimination. Similar findings were obtained by Hicks (1958) and Davis (1984), who trained subjects (rhesus monkeys and a racoon,

respectively) to discriminate three items from 1, 2, 4 and 5 items. However, unlike the Davis and Bradford study pattern recognition may account for both of their results.

In the third phase of Davis and Bradford's experiment the experimenter was absent from the room during trials and was thus unable to deliver punishment if the rat ate more than its target number, or to deliver reinforcement on correct trials. By the fourth trial only one rat was restricting its food intake. When punishment was reintroduced from the next room performance returned to phase two levels within two sessions, suggesting that motivational factors were responsible for the deterioration in performance.

However, there is an alternative explanation for these results. The authors did not measure the time that subjects took to consume their designated number of food items. Perhaps subjects ate for a relatively fixed period of time. This would explain why performance was initially disrupted by the transition from pellets to sunflower seeds, as the same number of sunflower seeds may have taken more or less time to consume than the pellets. Varying time to eat, perhaps by using different sized food items, may act as a control to rule out temporal cues.

Watanabe (1998) trained pigeons to peck at four objects and to withhold responding to two objects. Training involved discriminating "four" versus "two", using red balls, green balls, screw nuts, a mixture of red balls and screw nuts and collections of various objects (such as stones and twigs). Once performance under the training procedure reached 80% correct three types of tests were introduced. The first test was a generalization test, where one to five balls, screw nuts or different objects were presented. This test aimed to determine whether subjects were responding on the basis of absolute or relative number. If relative numerosity determined

discriminations then subjects would respond to five objects and not four when presented with a five-object four-object pair. The second test involved the presentation of large and small balls. Two large balls take up more area than four small ones, so if discrimination is controlled by area then subjects would respond to the two ball set and not the four ball set. The third test examined whether heterogeneous objects (red balls and screw nuts) would be seen as belonging to the same set.

In general, subjects responded more to four objects than to five objects. Subject's discrimination was also maintained when different-sized balls were used, and when heterogeneous objects were used. These results suggest that pigeons are able to discriminate numbers such as 'four' and 'two' and that these discriminations are based on the absolute properties of stimulus number and not a simple more versus less relation. Their finding is consistent with results from Davis and Albert (1986) who trained rats to respond to three noise bursts but not to two or four noise bursts.

Despite difficulties with alternative explanations for some of the above results, there is some suggestion that animals may be taught to discriminate on the basis of the absolute number of items in a set, as opposed to the relative number of items in that set.

Summation, Ordinality and Transitive Inference:

Ordinality is one necessary feature of counting, and demonstrating its existence in animal species would suggest that counting might be within their abilities. Olthof, Iden and Roberts (1997) examined squirrel monkeys' ability to learn about the ordinal relationship between Arabic numerals. Their subjects were two nine-year old monkeys named Jake and Elwood.

In Experiment 1, arabic numerals were painted onto wood lids that sat over food wells containing a number of peanuts. The number of peanuts in the well corresponded to the number painted on the lid. Two numbers were presented at a time and Jake and Elwood were required to choose between them. When they touched one of the lids, the other food well was removed and the monkey was allowed to eat the peanuts in the container of their choice. Training consisted of five phases throughout which the numbers 0, 1, 3, 5, 7, and 9 were successively introduced. Subjects were trained with most of the possible number pairs, although some pairs were withheld for a subsequent test. In each phase they had to reach 80% correct before the next phase was introduced. In all phases, correct choices were defined as the largest number presented.

Once training had been conducted with the selected number pairs, pairs that had been withheld were then presented interspersed with the training pairs. On such novel pairs Jake and Elwood's performance was between 90-100% correct over ten sessions. The high degree of accuracy achieved with these novel pairs suggests that Jake and Elwood had learned about the ordinal relationship between number symbols.

Jake and Elwood were also tested for their ability to choose the largest number from a pair of sums (each item in the pair was made of two component numbers). The numbers 0, 1, 3, 5, 7 and 9 were again used, and two-number versus two-number parings were presented together. This required the monkeys to sum the two numbers and compare that to the sum of the other two numbers. Both Jake and Elwood were accurate at this task right from the first trial. Furthermore the two numbers often summed to a number that they had never been trained on, i.e. $3 + 3 = 6$. Also, for example, $5 + 7 = 12$ exceeded the range of numbers on which they had been trained. Nevertheless, they nearly always chose the larger number. Number pairs were also

presented to rule out the possibility that Jake and Elwood were using a rule such as 'choose the card with the largest single number' instead of summing the numbers and comparing the sums. This type of rule would produce an incorrect choice in cases such as 9 + 0 versus 7 + 5; however Jake and Elwoods' performance remained at about 75% on these trials.

The results of these experiments suggest that Jake and Elwood understood the ordinal relationship between six arabic number symbols. That they were accurately able to choose the largest number or sum from a pair also suggests that the monkeys represented the quantities of food associated with each number in some way. The authors identified several possible mechanisms that may account for these results. First, the number of peanuts associated with each number could be stored as a single volume or mass. This would be a non-numerical representation. Second, the monkeys may have developed a prototypical image of, e.g., "5 peanuts". In this case different spatial configurations of the same number of peanuts must be transformed to one image. Third, each peanut encountered may produce a neural pulse (e.g. Meck and Church, 1983) which accumulates so that, for example, the number 5 is associated with 5 pulses. This is an example of a non-verbal counting system. However, they did not speculate which of these three possible mechanisms the subjects actually used.

In a similar experiment by Rumbaugh, Savage-Rumbaugh and Hegel (1987) subjects also summed food wells containing chocolate. However, in this experiment the chocolate pieces were a constant size so that quantity was perfectly correlated with number. Thus, performance may have been the result of volume discrimination rather than numerical discrimination. The similar performance of subjects in the two experiments may suggest that subjects in Olthof et al's experiment were also discriminating on the basis of volume.

Other researchers have also found evidence of ordinal representations in animals. Brannon and Terrace (2000) trained rhesus monkeys to order stimulus sets containing one to four items in either ascending or descending order. The monkeys were presented with four square stimuli on a video screen. Each stimulus was a constant size and had between one and four elements that varied in size, shape and colour. The subjects were required to touch each stimulus in the correct order: 1, 2, 3 and 4 elements for the ascending series, and 4, 3, 2 and 1 elements for the descending series. During transfer tests with new stimuli, subjects performed as accurately as they did with the training stimuli on the first trial. This suggests that subjects were not simply memorizing stimulus sets to produce the correct response.

Another explanation is that subjects assigned each numerical stimulus to a nominal category, i.e. A B C D, and then responded to these categories in an arbitrary order. To show that subjects had in fact learned an ordinal numerical rule, not a nominal one, novel numerosities 5-9 were introduced. Values 1-9 were paired in all possible combinations giving familiar-familiar pairs, familiar-novel pairs and novel-novel pairs. The two monkeys that had learned the ascending series 1-4 correctly ordered the novel numerosities at levels well above chance from the first trial. However, the monkey that learned the descending series did not perform much better than chance on novel-novel comparisons.

These results indicate that monkeys are able to represent the numerosities 1-9 on an ordinal scale. How they do this is yet to be determined. Subsitising and counting are two explanations, however subsitising does not appear to account for these results as numerosities as large as 9 were used. Humans are unable to subsitise beyond 4-5 and it thus would seem unlikely that monkeys can. Although these results do not constitute direct evidence of counting, ordinality is one requisite for counting

ability. The demonstration that monkeys are capable of ordinal representation of numbers suggests that counting may be within the grasp of this species.

Ordinal knowledge can be inferred from the demonstration of counting, however, it is also possible to establish ordinal knowledge independently from counting ability. Logical reasoning is strongly related to mathematical ability, and in this capacity the ability to make transitive inferences may be used as an indicator of ordinal representation. Transitive inferences are inferences made about the relationship between two premises when no explicit information has been provided regarding that relationship. For example, if A is bigger than B, and B is bigger than C, then A is bigger than C. No information was given regarding the size relationship between A and C, and yet it is still possible to infer that A is bigger than C. Reaching such a conclusion denotes transitive inference ability, and suggests that the individual is able to represent values on an ordinal scale. Piaget (1955) found that children could not perform transitive inferences until they were approximately seven years of age. However, subsequent research has indicated that children far younger than seven can perform these tasks given careful experimental designs (e.g. Bryant and Trabasso, 1971). This raises the possibility than non-human animals may also be able to perform transitive inferences.

Gillian (1981) tested chimpanzees on transitive inference tasks using a method developed by McGonigle and Chalmers (1977). A series of coloured cubes, denoted A to E, were successively paired, with only one of the pair associated with food reinforcement, giving; A-B+, B-C+, C-D+ and D-E+. Five item series were used since in a three item series the reinforcement histories of A, which would always be baited,

and C, which would never be baited, specify which of the pair to choose. Using a five item series the untrained pair BD can be presented. Both B and D have equally been associated with reinforcement and non-reinforcement, therefore to decide between them the subject must have placed these items on an ordinal scale with the other items in the series. Of Gillian's three chimpanzees Sadie scored 12 from 12 on the BD comparison, while Jessie and Luvie scored 7 and 5 from 12 respectively. These results indicate that Sadie and Jessie were able to correctly infer the relationship between B and D and suggest that they represented the items A to E on an ordinal scale.

Von Fersen, Wynne, Delius and Staddon (1991) also examined the ability of pigeons to make transitive inferences. A five item series was again used, with arbitrarily shaped black and white stimuli projected onto two response keys. Subjects were trained with A+B-, B+C-, C+D- and D+E- pairs. The test pair BD resulted in performance ranging between 75-100% correct. The consistent choice of B over D on test trials with untrained pairs suggests that pigeons are as capable as primates of making transitive inferences.

Davis (1992) used a similar design to test transitive inferences in rats. Instead of visual cues Davis used olfactory ones, specifically two tunnels each baited with one of five distinctive odors. In any trial only one tunnel could be accessed to receive food reinforcement and this was designated the correct choice. Trained pair comparisons included A-B+, B-C+, A-C+, C-D+, D-E+, C-E+ and A-E+, and the test pair was BD. Three of four subjects met criterion during training with each of the seven comparisons. During testing with the untrained BD pair one subject correctly chose D in all presentations while the other two subjects correctly chose D on 14 of 16, and 15 of 18 comparisons. These results are similar to those of Gillian and further suggest that a range of species, including rats, are capable of making transitive inferences.

All three experimenters also examined the effect of transforming the series from a linear series to a circular one. The premise E<F was added and then subjects were trained with an inconsistent A+F- pair, thus disrupting the logical structure of the series. All found that performance from previous BD tests was disrupted, with either a slight reversal in preference or no preference for either item in the pair. This finding indicates that making a series circular disrupts subject's ability to make transitive inferences by eliminating the linear order of the series. This further suggests that primates, rats and pigeons represent items on an ordinal scale.

Counting:

Experimenters have found evidence suggestive of counting in several species including chimpanzees, rats and parrots. As previously mentioned, in order for an individual to count several conditions must be fulfilled. There must be a one to one correspondence between the items to be counted and the series of tags used. These tags must be applied in a stable order and there must be a recognition that the final tag applied represents the numerosity of the set. These tags do not necessarily have to be verbal; they may be behavioural or even mental.

Matsuzawa (1985) has examined enumerative ability in chimpanzees. Ai, a female chimpanzee, had previously been trained to name 14 objects and 11 colours by choosing among a set of symbols that represented each. Five forms and five colours were selected to form sample stimuli for numerical training. Sample items were presented in the display window and numeric keys were lit. Immediate feedback was produced via different noises following correct and incorrect choices. Food reinforcement was only provided following a number of consecutive correct trials.

Training began with the presentation of only one or two red pencils and the corresponding number keys. Gradually the number of items and the number of numerical keys were increased from one to six. Generalisation tests using new objects and colours were introduced following the introduction of each new number. Accuracy of trained samples on probe trials ranged from 0.97 to 0.99, and on generalisation trials from 0.33 to 0.87. These results are consistent with those of Ferster (1967) who taught chimpanzees to 'write' the number of a set of items in binary form. In a second test, Ai was required to name the number, type of object and colour of the items in the display. While number naming was always less accurate than either object or colour naming overall accuracy was still high, and Ai generally took less than three seconds to choose the colour, object and number of the sample. Again, although these results, and those of Ferster, are consistent with counting, it has not been demonstrated that subjects represented items on an ordinal scale.

Capaldi and Miller (1988) have concluded that not only can rats count but that they do so quite naturally. In their experiments rats were required to count the number of successively presented food reinforcers rather than geometric figures or auditory stimuli. They argue that animals are disposed to count reinforcing events in their natural environment, consequently using biologically relevant stimuli provides a fairer test of counting ability than would the stimuli that are normally used.

Experiments 1 and 2 involved the multiple presentation of two different series of runway trials. The first series consisted of two reinforced trials followed by one non-reinforced trial (RRN); the second series consisted of one non-reinforced trial then two reinforced trials followed by another non-reinforced trial (NRRN). Once the rats reached the goal box they were confined for 60 seconds on non-reinforced trials

and 15 seconds on all other trials. In this procedure rats ran more slowly on the non-reinforced trials than they did on the reinforced ones. In order for them to anticipate the upcoming non-reinforced trial, they must either count the number of R events (i.e. two R's means no reinforcement) or sum the confinement times of the R events (i.e. 30-seconds means no reinforcement). If summing confinement times to reinforcement (CTR) is the method by which rats anticipate the non-reinforced trial, then disrupting this interval should also disrupt subsequent running times. In Experiment 1 reinforced trial confinement times were increased to 30 seconds, while in Experiment 2 they were decreased to 7 seconds. Neither change had any effect on responding, suggesting that subjects were counting the number of R trials. The authors concluded that counting was a primary process since CTR was an equally valid cue that was not utilised.

In Experiment 3 subjects were divided into two groups. The first group was trained on the three-trial series RRN and a single N trial, presented in an irregular order. The second group was trained on the four-trial series NRRN and a single R trial, also in an irregular order. During training, anticipation of N could be based on several number and timing cues, that is, trials, responses and confinement times. During test sessions RRN and NRRN series were presented in an irregular order. For the first few presentations of these trials, the only reliable cues for the occurrence of N were the number of R events and the sum of confinement times for R events. As in Experiments 1 and 2, rats ran significantly slower on the N trials than they did on the R trials. This suggests that despite the availability of alternative cues during training sessions, rats learned about the relationship between R and N trials on the basis of number. Another finding was that rats tended to run more slowly on the terminal N trial of the longer sequence. The authors proposed that the rats might have learned (from both series) that non-reinforcement is more likely later in a series than earlier.

Despite their demonstration that timing cues were not used, the results of Capaldi and Miller's experiments in no way demonstrate that distinct numerical tags were applied to the R trials, and thus that counting was occurring.

An alternative explanation for the results in Experiment 3 is that the subjects simply learned to run slowly if the previous two trials were the same and to run fast if they were different. In Experiment 4 the two series from Experiment 3 were changed to RRRN and NRRRN. If the above proposition is correct, then the rule "two the same-go slow, two different-go fast" would result in the third R trial in the series being run slowly in initial trials. This was not the case, indicating that subjects did not base their discriminations on such a rule learned in the first two experiments.

For counting to be a primary process the number of events must be used in spite of the presence of other valid cues. In Experiment 5 rats were given the opportunity to use either the "slow-if similar and fast-if different" strategy or to count the number of R trials. In order to count, in the initial phase of the experiment rats had to categorise corn pops (R') and rat pellets (R) as different (corn pops vs. pellets) or similar (food items). Rats received multiple presentations of the four trial series R'RRN and a single N trial. Rats were then shifted to either the series RRN and NRRN or to the series RRRN and NRRRN. If the subjects were not counting the number of reinforcing events they would be unable to master the second series pair, since it violates the "slow-if two the same and fast-if two differ" rule. However, if during the training phase the subjects learned that two successive pellet trials (RR) and three successive food events (R'RR) both signalled non-reinforcement, then both series pairs should be mastered. Throughout training running time for the N trial in the sequence increased. When the series was shifted to the two new series pairs, those rats exposed to the latter pair did show anticipation of N trials, as evidenced by slower

running times. This indicates that they did not use the simpler two-same versus two-different rule in training sessions.

Experiment 6 attempted to demonstrate the application of abstract number tags to corresponding events by showing that different types and quantities of reinforcers (corn pops or pellets) were given an abstract tag to predict the non-reinforced trial. In this experiment rats received one corn pop (P) and six pellets (P*). Half of this group was trained on P*PPN and PPN series pairs and the other half on PP*P*N and P*P*N pair; after 16 days of initial training these conditions were reversed. Both of these series pairs can be represented as R'RRN and RRN. After training, all rats were shifted first to R'RRN and R'R'N trials and then to RRN and NRRN trials, where confinement was reduced from 15 seconds to 7 seconds, and R represented pellets. Slower running on the N trials after the shift from training was apparent on the first day of the test. This finding suggests that the rats applied abstract number tags to events, irrespective of the absolute properties of the trials. Thus, they recognised that the tag for one event (either R or R') during training signalled R, and that the tag for two events (either RR or R'R') signalled N. Rats had to recognise that R'R' and RR both signalled the same event, irrespective of their concrete properties (1 corn pop vs. 6 pellets). What they had in common was an abstract number tag corresponding to two events. When rats were shifted to the second series pair, with reduced confinement time, responding was not disrupted, again demonstrating that summing of the CTR's over trials did not account for the results.

Experiment 7 demonstrated generalisation of counting to new reinforcing events. Four different reinforcing events were presented: rat pellets (R), corn pops (R'), honey snacks (A) and cocoa puffs (B). As in Experiment 6, initial training was conducted with the series pairs R'RRN and RRN and with RR'R'N and R'R'N. Once

discrimination responding was established rats were shifted to new series pairs involving the new reinforcers (ABBN / BBN and BAAN / AAN) to test whether performance would generalise. In the first few days of the shift, slow running behaviour in the terminal N trial was only slightly disrupted, supporting the results of Experiment 6 and suggesting that new reinforcers were allocated the same abstract tags as reinforcers used in training.

The authors proposed that not only were rats applying abstract numerical tags to the reinforcing events, they were also applying tags in a set order, irrespective of the order in which the different reinforcers were presented. Over training and test trials the four reinforcers were presented in two positions each. One position (R, R', A or B alone) signalled the occurrence of the second position. The second position (RR, R'R', AA or BB) signalled the non-reinforced trial. Each position was assigned a numerical tag that did not depend on the reinforcer that occupied that position.

In summary, Experiments 1-3 suggest that rats based their discriminations on the number of reinforcing events and not time spent confined after each trial. Experiments 4 and 5 similarly demonstrated that the number of reinforcing events and not simply a "slow-if-same and fast-if different" strategy was used to anticipate non-reinforced trials. In Experiments 6 and 7 it was demonstrated that numerical tags were being applied to the positions containing one and two reinforcing items, and that this performance generalised to new reinforcing items. The authors concluded from their results that when accurate numerical cues are available animals count reinforcing events routinely. This implies that counting will only be abandoned when numerical cues are unpredictable.

Sarah Boysen (1989, 1992) has published several fascinating experiments examining numerical ability in chimpanzees. One chimpanzee, Sheba, was trained to recognise that arabic numerals correspond to different numbers of objects. She was not only able to choose the arabic numeral that corresponded to the number of items in an array, but also to select the correct number of items when presented with an arabic numeral. Sheba's repertoire included arabic numerals from one to eight.

Three main findings have emerged from these experiments. Sheba has demonstrated partitioning of items, transitive inference and an ability to perform operations on numbers. Firstly, when the number four was introduced (directly as an Arabic numeral, rather than corresponding to items in an array), Sheba began to show evidence of motor tagging. She would point at, touch or move items to be counted until the correct numeral was chosen. It was hypothesised that as her repertoire increased motor tags may have helped her keep track of the larger numbers. This finding indicates that she was partitioning those items that had been counted from those yet to be counted.

As part of Sheba's ongoing training Boysen (1992) tested whether she was able to make transitive inferences. Five pairs of coloured boxes that formed an ascending series ABCDE were presented. In training only one box in the pair was ever baited with food i.e. A-B+, B-C+, C-D+ etc. On blind trials the non-adjacent pair BD were presented. Sheba consistently selected the baited box from the pair. Next the arabic symbols one to five were used instead of colours. From the test pair 2/4 Sheba always selected 4. This suggests that she recognised the ordinal relationship between the numbers one through to five.

Boysen and Berntson (1989) also examined Sheba's ability to perform operations on numbers. In an experiment designed to test whether Sheba could

perform addition, 0-4 oranges were placed in two of three separate sites. Sheba had to move among the sites, return to the work area and select the arabic numeral that corresponded to the total number of oranges at all of the sites. Remarkably, Sheba was able to correctly sum the oranges right from the first trial. In the next stage oranges were replaced with cards displaying Arabic numerals. Sheba readily generalised her experience to this new task, demonstrating transfer of ability. This demonstration of summation differs from that of Rumbaugh et al. (1987) in that perceptual fusing or subsitising of the items to be summed does not appear to be a possible explanation for Sheba's performance. Because the oranges and then arabic numerals were separated in space and time, perceptual representations would have to be stored and then added together once all three sites had been encountered. This goes far beyond the features proposed by Davis and Perusse (1988), and does not appear to be a possible explanation for counting.

In a related experiment Sheba also demonstrated subtraction ability. Sheba was seated in front of an array of up to four oranges. Once she had looked at the array it was covered up and the experimenter removed oranges one at a time. Each of the oranges was shown to Sheba as it was removed and was then placed out of her sight. She was then asked how many were left and was required to point to the arabic number card corresponding to the number of oranges remaining. Sheba's performance on these trials, where between one and four oranges were removed, was about 85% correct. These two results seem to demonstrate that Sheba was able to perform simple mathematical operations on a limited range of numbers. Davis and Perusse (1988) noted that the range of numbers that Sheba could use were within the range for subsitising. However, subsitising is characterised as a perceptual process and Sheba's

apparent ability to subtract and add suggests some form of cognitive process. Thus, subsitising is an unlikely explanation for Boysen and Berntson's results.

Overall the work done with Sheba constitutes strong evidence that chimpanzees are able to count. Of the requirements that must be met for counting; including cardinality, ordinality, transfer of ability to new stimuli and a partitioning of items, all have been demonstrated in experiments with Sheba. It is thus seems probable that Sheba, as well as other chimpanzees, possess the ability to count.

Irene Pepperberg (1994) has been working with Alex, an African gray parrot, on interspecies communication and animal cognition since 1977. Alex can vocally label collections of two to six homogeneous and heterogeneous objects, however the possibility that he may be using some form of non-enumerative process, such as subsitising, has not yet been ruled out. Work with humans has suggested that the presence of distracters in a test may distinguish between perceptual processing and counting. When tests require the subject to distinguish between several features of an array, such as colour and form, subsitising does not appear to occur. Requiring Alex to discriminate the number of a certain item, i.e. red keys, among other coloured keys and other red objects may illustrate his ability to count.

During training four different quantities of two different colours and two different objects were presented on a covered tray. After allowing Alex to touch each object with his tongue they were randomly scattered on the tray. When Alex responded correctly he received praise and was given access to the objects referred to in the question (or allowed an alternative request). If incorrect, the examiner turned their head and emphatically said NO, the trial was then repeated (correction procedure). Accuracy on all first trials was 83.3% and overall accuracy (first-trial and

correction trials) was 83.1%. One point worth noting is that Alex was not more accurate with smaller numbers than he was with larger numbers, as might be expected if he was subsitising. Latency to respond is also often used to differentiate between subsitising and counting in the human literature, where subsitising is thought to be a faster process. However, Alex's readiness to respond was related to his interest in obtaining the objects in the sample rather than to any other factor. Another point of interest is that eight of nine errors involved producing the correct number of an incorrect subset, suggesting that these errors were failures of stimulus control by the request and not in his counting ability.

These results demonstrate that Alex is able to accurately label quantities from heterogeneous arrays. However, while Alex's performance is consistent with counting these results, like those of Ferster and Matsuzawa, do not demonstrate that Alex possesses ordinality. Thus according to Davis and Perusse Alex's ability would be termed protocounting, for lack of conclusive evidence of counting.

While the work of Ferster (1957) and Matsuzawa (1985) lack demonstrations of ordinality, other work (Brannon et al., 2000; Gillian, 1981) suggest that chimpanzees do possess, or can be taught, ordinal knowledge regarding numbers. The work by Von Fersen et al. (1991) and Davis (1992) further suggests that other species can also make ordinal judgements and thus may be capable of counting.

Pigeons and rats are both very accurate at timing sequences and events. Therefore, any work examining their numerical ability must ensure that timing is not the process on which subjects base their discriminations.

Fernandes and Church (1982) controlled for temporal cues in a bar pressing task with rats in order to demonstrate that number and not time may be used in discrimination tasks. Rats were required to discriminate between sequences consisting of either 2 or 4 sounds. Varying the Inter-Stimulus-Interval (ISI), which also varied the total sequence duration, controlled for sequence density. Duration of the sounds were controlled by providing two versions of the 2-sound sequence, one that consisted of two short sounds and the other of two long sounds. The two-long sound sequences were of the same duration as the 4-sound sequences, forcing correct discriminations between the two to be on the basis of number.

Rats classified the 2-short sequences more accurately than either the 2-long or 4-short during training, however during tests sequences of four-sounds were classified as 'many' significantly more than were either of the two-sound sequences. Moreover, there was no significant relationship between the proportion of 'many' responses and the ISI's. From these results the authors concluded that when sound duration, sequence duration and the interval between each sound are controlled for rats are able to discriminate between numbers of sounds. Thus, it appears that rats do not necessarily base judgements on the duration of events, but are capable of making numerical discriminations.

Fetterman (1993) undertook an evaluation of the joint contribution of time-based and number-based cues to discrimination on a numerosity task. Pigeons responded on a symbolic matching to sample procedure. When the centre key was lit half the trials consisted of an FR 10 schedule and the other half an FR 30 schedule. Once the FR requirement was completed the centre key was darkened and the side keys were lit red and green. For two birds a response to the green key was correct following the smaller ratio requirement and a response to the red key for the larger ratio requirement. This was reversed for the other two birds. Probe trials were inserted during testing whereby the FR value was intermediate to the small and large values. Time taken to complete each ratio requirement was recorded, and formed the basis for the time-based task. In these tasks, responses to the choice keys were reinforced after a duration had elapsed that was equal to the median time taken to emit the FR 10 requirement in the number-based task. The number of pecks to the centre key in the initial stage was irrelevant to reinforcement. Probe trials were also inserted in this test, where the interval duration required for reinforcement was equal to the median time for the intermediate probe requirement to be completed in the number-based task.

Examination of the times taken to emit responses under the large and small FR requirements (as well as probe trials) indicated that time influenced the bird's choices over and above the number of responses. The authors used multiple regression analyses to assess the control of time and number on choice. For two birds time exerted the major control on choice, for another bird number exerted the major control. For two other birds the results were less consistent. Fetterman concluded from this that both time and number dimensions were important in subject's choice discriminations, despite time being a less accurate cue. He further suggested that

multiple sources of information might be useful to predict important events when the sources of information are less than perfectly correlated.

Breukelaar and Dalrymple-Alford (1998) proposed that even when event duration, sequence duration and the interval between events are controlled, temporal ratios still co-vary with number. When a fixed number of events occur periodically the ratios of the onset interval, event duration and inter-event duration to total sequence duration are constant. Consequently, these ratios vary with the number of events in a sequence and are possible confounds.

In their first experiment, Breukelaar and Dalrymple-Alford trained rats to choose one of two levers in response to two sequences of sounds where both number and duration were equally good predictors of reinforcement. During training rats were required to discriminate between a sequence of eight sounds and a sequence of two sounds. In test trials the number of sounds either varied between two and six with sound duration held constant at four seconds, or the number of sounds was held constant at four while sound duration varied between two and six seconds. When number was held constant the rats accurately discriminated on the basis of time. However, when duration was held constant rats ignored the numerical cues and continued to discriminate on the basis of time.

In Experiment 2 rats were trained with separate time-relevant or number-relevant cues. Subjects were then tested with probe signals to determine which cue was predominantly used. For the number relevant trials during training stimulus duration was held constant at four seconds and number was either two or eight sounds. For the time-relevant trials, time varied between two and eight seconds while number was held constant at four sounds. In the test trials number cues and duration cues

varied between three and six sounds and three and six seconds respectively. Subjects' performance was as accurate for temporal discrimination as it was in Experiment 1, while numerical discrimination gradually improved over 15 days. Thus, rats were capable of learning numerical discriminations when they were explicitly taught them. However, performance under numerical cues was significantly poorer than that for temporal cues, suggesting that timing is an easier task than enumeration.

Another possible explanation for this finding is that the rats learnt the temporal discriminations first and were consequently more disposed to use them. Experiment 3 examined this possibility by simultaneously training naïve rats to make separate time and number discriminations. The same procedure as in Experiment 2 was used for training and testing. The authors found, as for Experiment 2, that temporally based discriminations were significantly better than numerically based ones. These results indicate that rats are less sensitive to changes in numerosity than changes in timing. However, the fact that subjects could learn the numerical discriminations, albeit with difficulty, illustrates that regularity of timing is not necessary for number discrimination by rats.

Davis and Memmott (1982) proposed that animals use numerical cues as a last resort when making decisions. On the basis of this it would be predicted that even when animals have been trained to use numerical cues, they will not do so unless there are no other reliable cues to reinforcement. In Experiment 4 rats were provided with ambiguous temporal and numerical cues in which the two cues indicated that opposite responses were correct. If rats are disposed to use number and time equally, choices should be fairly evenly distributed between the two levers. When both temporal and numerical cues were ambiguous, responding appeared to be entirely in the control of time (i.e. 2 events/ 8 seconds and 8 events/ 2 seconds). Even when temporal cues

were inaccurate they appeared to influence responding in the presence of accurate numerical cues. Only in instances where temporal cues were neutral did number guide discrimination (i.e. 2 events/ 4 seconds and 8 events/ 4 seconds).

These results indicate that although rats are capable of learning and using numerical discriminations they only do so when other, more salient, cues are not available. This supports Davis and Memmott' (1982) proposition that rats use numerical discrimination as a last resort.

Roberts and Boisvert (1998) have also examined the roles of time and number in discrimination trials with pigeons. They note that the peak procedure has been used extensively to study timing but has not been applied to counting. When animals are required to respond according to a FI schedule response rates are usually low over the first part of the interval followed by rapidly accelerating responding over the latter part. The peak procedure has revealed that w-hen non-reinforced probe trials, far longer than the original FI, are inserted responding generally peaks at the FI value. Roberts and Boisvert proposed that if the same process is responsible for timing and counting then when animals are reinforced for responding to time or number the peak procedure should produce similar results.

In their first experiment, birds were reinforced for the first response after 20 one-second red key flashes had occurred. It was expected that under the peak procedure pigeons would show elevated responding near 20-seconds and 20 flashes. The addition of trials where the speed of the flashes varies should indicate whether pigeons use time or number to make discriminations. Flash rate was varied so that flashes occurred at the training rate (medium rate), two per second (fast rate) or one every two seconds (slow rate). The medium and slow flash rates peaked at the same

time while the fast rate peaked slightly sooner. This suggests that the birds used time or number, depending on which cue crossed a threshold for terminating responding first.

Experiment 2 examined whether either time or number would control responding if only that cue predicted reinforcement. Subjects were either reinforced after a FI of 20-seconds or a FR of 20 flashes. Thus, depending on the flash rate the number and time that produced reinforcement varied. For birds that were time-reinforced, peaks occurred at 20 seconds for all three flash rates, indicating that they accurately timed. For number-reinforced birds peaks shifted along the time dimension, indicating that these birds also accurately used number.

These findings suggest that selective reinforcement can produce either process, and that birds will use either cue. However, the fact that half of the birds were selectively reinforced for using number and not time could suggest that previous reinforcement produced a tendency to continue to use number when it may not be a primary process. It has already been demonstrated that rats can use number, but also that they do not generally choose to (Breukelaar and Dalrymple-Alford, 1998). The same might also be true of pigeons.

It is also possible that the pigeons were sensitive to the flash rates and, for example, learned that at the faster rate, reinforcement occurred after 10 seconds. If this was the case then the pigeons were in fact timing as opposed to using number discrimination. In Experiment 3 the medium flash rate for the number-reinforced pigeons remained the same. The fast and slow flashes began at their previous rate but changed to the medium rate half way through. Thus, the fast-medium flashes were 15 seconds in duration (as compared to 10 seconds previously) and the slow-medium flashes were 25 seconds in duration (compared with 30 seconds). If time and not

number was controlling responding, then response peaks for fast- and slow-medium flash rates should have shifted from their previous points to the right and left respectively. This was not the case, and birds' responding continued to peak at the original values of 10 and 30, indicating that number and not time was controlling the pigeons responding.

Experiment 4 examined whether pigeons would continue to respond according to time or number when new time and number values were introduced and differential reinforcement was not provided. If pigeons previously reinforced for responding to number began to respond according to time this would suggest that timing is a primary process, or vice versa. In fact pigeons continued to respond according to time or number cues with the new values. This suggests that the differential reinforcement given in Experiments 2-4 biased the birds towards using only one cue dimension. These results contrast with those of Breukelaar and Dalrymple-Alford (1998) who concluded that for rats time, and not numerosity, was the primary discriminative process.

These results illustrate two things. With selective reinforcement both pigeons and rats are capable making discriminations on the basis of number. However, there is some suggestion that temporal cues will be used over numerical cues, even when temporal cues are inaccurate. Also, numerical discrimination appears to be poorer than temporal discrimination. This indicates that given a choice pigeons and rats will time rather than use number unless they are explicitly forced to do otherwise. The implications of this for research on numerical competence are that temporal cues must be controlled if results are to reflect numerical ability.

Mode Control Model of Timing and Counting

Gallistel and Gelman (1992) proposed that both human and non-human animals have a pre-verbal counting mechanism. Of interest is how such a mechanism might work. Meck and Church (1983) developed a mode control model of counting and timing processes in rats. During training rats were reinforced for a right response following a two-cycle noise of 2-seconds duration, and a left response following an eight-cycle noise of 8-sec duration. On additional test trials, control by number and time was assessed by holding one dimension constant at 4-seconds or 4-flashes, and varying the other. Thus, during training both time and number were valid cues for reinforcement, while during testing time and number were confounded so that on any one trial only one dimension was a relevant cue to reinforcement. Meck and Church found that rats could correctly respond according to either cue, and that accuracy was equivalent for time and number discriminations.

From these results Meck and Church proposed that subjects were timing and counting simultaneously, and that the same mechanism is used for both processes. They hypothesized that an internal mechanism produces pulses that are controlled in several modes. Figure 1 illustrates these modes.

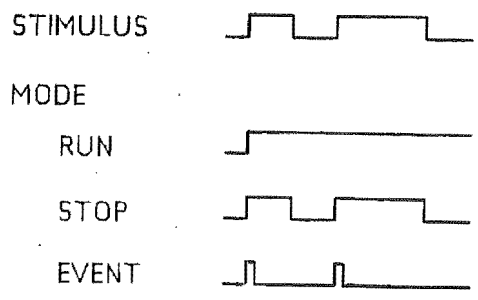

Figure 1. Three modes of operation of the accumulation process, from Meck and Church (1983)

In the run mode a pulse starts with the presentation of the stimulus and continues until the end of the trial. In the stop mode a pulse is emitted whenever the stimulus occurs, and continues for the duration of the stimulus. In the event mode the onset of the stimulus produces a pulse of (relatively) fixed duration, independent of stimulus duration, so that for each presentation of the stimulus one pulse accumulates. The run and stop modes are used for the estimation of time, and the event mode for the estimation of number. Figure 2 illustrates how subjects use these modes to guide behaviour.

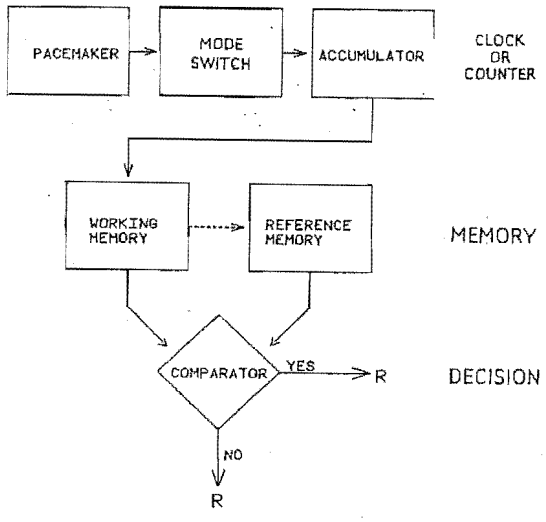

Figure 2. Mode control model of timing and counting, from Meck and Church (1983)

The pacemaker produces pulses and the mode switch closes to pass pulses to the accumulator. This system may act as either a clock or a counter depending on which mode is being utilised. In either case the value in the accumulator is then passed into working memory. A remembered accumulator value from a previous response at the time of reinforcement is stored in reference memory. The current accumulator value in working memory is then compared to the value in reference memory, and a decision about responding is made.

This model assumes that on any given trial the pacemaker emits pulses at a fixed inter-pulse interval (IPI), and that across trials this IPI is normally distributed. With experience the animal learns the accumulator value that is associated with reinforcement. This value is stored in reference memory and is used as the comparator value. If the current accumulator is close to the reference value for reinforced left responses then the animal responds left, if the value is closer to the reference value for reinforced right responses the animal responds right.

Support for the idea that timing and counting are controlled by the same mechanism comes from trials using methamphetamine. It has been demonstrated that methamphetamine speeds up the internal clock and produces a shift in response functions for time. Meck and Church (1983) found that methamphetamine produced a 10% shift in the psychophysical functions for number as well as duration. Furthermore, if the same internal pacemaker is used for counting and timing it should be possible to determine the quantitative measure of equivalence between an increment of one count and one unit of time. Meck and Church determined that one count and one unit of time are equal to approximately 200 milliseconds, a finding that has been replicated by Meck, Church and Gibbon (1985). Because the pacemakers 'count' is somewhat variable the number of event pulses and the number of stimuli are not perfectly correlated. In human performance verbal tags may replace pacemaker pulses, removing variability and resulting in greater accuracy in counting.

Roberts and Mitchell (1994) sought to replicate Meck and Church's findings with pigeons. Pigeons were presented with sequences of light flashes that varied in both total duration and number of flashes. They found, as did Meck and Church, that pigeons processed both the temporal and numerical properties of the light flashes.

However, unlike Meck and Church, time exerted greater control over behaviour than number (a result consistent with the findings of Breukelaar and Dalrymple-Alford, 1998). They noted that according to Meck and Church's model when time and number indicate different responses there is no way to respond accurately. This occurs because the model does not make a distinction between timing and counting past the accumulator stage. Roberts and Mitchell tested subjects on sequences containing 8 flashes in 2 seconds and 2 flashes in 8 seconds, where time and number cues are ambiguous. They found that when counting cues were correct subjects could learn to respond accurately to these ambiguous sequences. This suggests that a counting contingency allowed them to select numerical information from working memory to the exclusion of temporal information. Initial performance was poor on these trials however, again indicating a tendency for pigeons to use temporal cues over numerical ones. Roberts and Mitchell also found that previous experience with numerical discriminations improved subject's performance on number tasks, as did Roberts and Boisvert (1998).

In a further experiment Roberts and Mitchell (1994) provided cues (different coloured key lights) after the flash sequence telling the bird whether to time or count. The addition of these cues improved performance, suggesting not only that pigeons process time and number simultaneously but that the coloured keys acted as cues for the selective retrieval of time and number information from working memory. This indicates that the two types of information are stored separately and can be accessed independently of one another. Consequently, Roberts and Mitchell have adapted the Meck and Church (1983) model to allow for independent representations of time and number at the accumulator stage. Figure 3 shows their modifications to the mode control model.

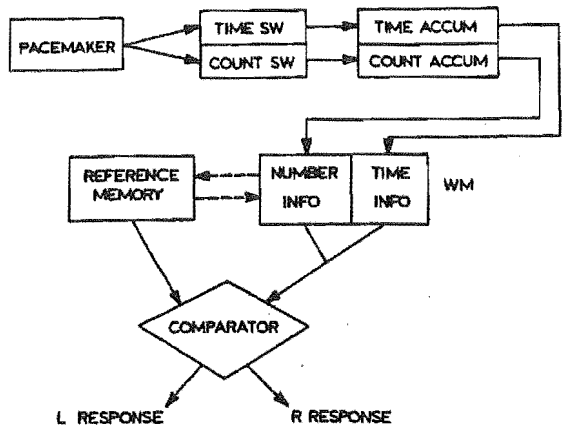

Figure 3. Mode control model of timing and counting adapted by Roberts and Mitchell (1994)

In this model time and number accumulators feed information into separate areas of working memory, allowing the two types of information to be accessed independently. The retrieval of one type of information over another appears to be influenced by past learning.

More evidence in support of Meck and Church's (1983) dual mode model of timing and counting comes from work by Roberts, Macuda and Broadbeck (1995, see also Roberts 1995, 1997). They noted that if timing and counting are controlled by the same mechanism, the manipulation of a variable that affects memory for time should also effect memory for number. A common finding in delayed matching to sample procedures is the 'choose small effect'. While memory for 2- and 8-second stimuli is equally good at 0-second delay, retention curves for the 8-second stimuli drop sharply as delays increase, resulting in a tendency to prefer the 'short' response key. Roberts et al. proposed that the accumulated pulse counts in working memory diminish over time. According the mode control model memory for number should diminish at a similar rate. In their experiment Roberts et al. held time constant and varied the

number of flashes presented. Pigeons were then required to match sample stimuli over varying delay periods. They found a clear choose small effect in number discriminations, indicating that pulses in working memory from the event mode were lost over time in a comparable fashion to pulses from the run mode. This finding further supports the view that timing and counting are controlled by the same internal mechanism.

The Present Research

A second use of number:

Most of the research on animal numerical competence has involved subjects making numerical discriminations about stimuli presented to them. However, there is a second way in which number is used. It is not only possible to respond discriminatively on the basis of numerical cues, but also for a response to be differentiated along a numerical dimension. This is analogous to the distinction between temporal discrimination and differentiation procedures (see Platt, 1979).

There is a small body of literature examining the production of responses in relation to numerical control of behaviour. Two types of response production procedures have been developed and have been termed response terminated counting schedules and stimulus terminated counting schedules (Brandon, 1981). In a response terminated counting procedure the subject reports whether it has reached the criterion number by switching to terminate the trial or by continuing to count. In stimulus terminated counting procedures during a choice phase the subject reports how many responses it emitted during an experimenter-terminated run.

Mechner's (1958) work on the internal cohesion of response runs is one example of a response-terminated procedure. In his experiment he examined the probability that a response run would be terminated as a function of the number of responses made in that run. Subjects were required to make N consecutive responses (response run) on lever A. On 50% of trials after the completion of lever A responding a further response was required on lever B. On trials where a response was not required on lever B reinforcement was delivered after the Nth response. On all other trials reinforcement was only delivered if the response to lever B followed at least N consecutive responses to lever A. N values were varied over trials so that either 4, 8, 12 or 16 responses were required on lever A. Mechner found that for each value of N the modal value for the termination of a response run was close to, although always exceeded, N. This indicates that subjects were able to terminate responding on lever A with reasonable accuracy. Subjects may have overestimated the number of responses required on A before switching because there was no upper limit for reinforcement on lever A responding, thus one or two responses over N was better than one or two responses below N.

Later work by Wilkie, Webster and Leader (1979) suggests that subject's ability to correctly produce the required number of responses in Mechner's procedure was not due to timing. Wilkie et al. modified Mechner's procedure so that each B response produced a blackout that was either of constant or variable duration. This reduced the likelihood that switching to A was controlled by the time between the first B response and the last B response. In all cases pigeons were required to peck key B twice before switching to key A (here A and B are reversed from in Mechner's study). The frequency with which different numbers of B key pecks preceded a peck to key A was examined. Subjects most frequently made two key pecks to key B before

switching to key A, and there were no differences between conditions where blackouts were of constant or variable duration. These results suggest that the number of key B pecks controlled switching to key A, rather than the time from the first key B peck to the last.

Rilling and McDiarmid (1965) used a stimulus terminated counting procedure to examine the discriminability of fixed ratio schedules by pigeons. Their procedure involved two steps. First subjects were required to peck the center key according to the fixed ratio (FR) schedule in effect, then the subject was required to peck one of two side keys to indicate which of the two schedules it had just completed. One fixed ratio schedule was held constant at FR50 while the other one was varied. Initially the second schedule began at FR35, the ratio was then increased by increments of 2 until performance fell below 60% correct. For both birds performance remained above 60% until the schedule reached FR47. These results indicate that pigeons are accurate at discriminating between two FR schedules whose values are close to one another.

Rilling (1967) sought to determine whether time or number was responsible for the discrimination seen in Rilling and McDiarmid's study. Subjects were required to discriminate between either two fixed interval (FI) schedules or two fixed ratio (FR) schedules in a procedure similar to that of Rilling and McDiarmid. The long FI was held at 45-seconds while the short FI was either 30- or 36-seconds. The long FR was held constant at 50 responses and the short FR was either 32 or 38 responses. In the FI schedules the number of responses emitted during an interval was systematically related to the accuracy of the discrimination between the two FI's, while duration of the center key or duration of the FI itself were not. In the FR schedules run time was

not systematically related to discrimination either, suggesting that the number of responses emitted, and not time, controlled discrimination.

Thus, work involving two different response production procedures has indicated that the number of responses emitted by a subject, and not simply the time taken to emit those responses, can act as discriminative stimuli guiding subjects behaviour. This illustrates that numerical competence procedures are not limited to the presentation of stimuli to subjects but that response production tasks can also be used to examine numerical ability.

The present experiment:

Several researchers have indicated that pigeons are capable of making relative numerosity discriminations, and there is some suggestion that pigeons can discriminate absolute, as opposed to relative, number (Wantanabe, 1998). The modification of Mechner's procedure by Wilkie et. al. further suggests that pigeons are sensitive to the absolute properties of number; however their demonstration was limited to a single number and may instead be due to stereotyped responding. Demonstrating that pigeons can accurately produce a range of numbers would further strengthen this finding, and may also suggest that pigeons can represent numerosities on an ordered continuum, which is one requisite for counting. Consequently, this demonstration may suggest that pigeons' are capable of more advanced numerical abilities such as counting.

Zeiler and Hoyert (1989) devised a procedure to examine temporal reproduction in pigeons. A signal key was illuminated for one of a range of durations.

After a short period with both keys darkened the centre key was illuminated with a FR15 requirement, after which the key was turned off. Reinforcement was only delivered if the FR15 was completed within the required range set by the signal duration. The lower limit of the range was the same as the signal duration while the upper limit was 1.5 times the signal duration. Thus, if the signal key was lit for 8-seconds subjects would only receive reinforcement if 15 responses to the centre key occurred within 8 to 12-seconds. Subjects were initially trained with pairs of signals, which were gradually increased until 10 different signal durations were used. Mean ratio times increased as signal duration increased, in a similar fashion to results with humans. However, mean ratio times always underestimated the duration of the signal. The authors concluded that large numbers of stimuli helped subjects match behaviour to the signal duration rather than treating each duration as an arbitrary discriminative stimulus.

The Zeiler and Hoyert study indicates that subjects can time their own behaviour to match a range of sample stimuli presented to them. Studies reviewed above (e.g. Fernandes and Church, 1982) have indicated that given the right controls subjects can be trained to use numerical information in the place of temporal information. This suggests that subjects may be able to reproduce particular numbers that have previously been presented to them. The current experiment combines a stimulus presentation procedure with a Mechner-like response production procedure in order to develop a new paradigm to study numerical competence in pigeons.

The advantage of this experiment over past response production procedures is that temporal cues can be controlled during the presentation phase to rule out temporal discrimination as an alternative explanation. Also, because the presentation phase

specifies the number of responses required in the production phase, a range of numbers can be used instead of a single number as in the Wilkie et al. study.

METHOD

Subjects

The subjects used in this experiment were four homing pigeons, numbered Q5-8, and were experimentally naïve. Birds were maintained at approximately 85% of their free-feeding weights by additional feeding, when necessary, after experimental sessions. Water and grit were continuously available in their home cages.

Apparatus

Experimental sessions were conducted in four identical chambers measuring 40cm by 40cm by 33cm. One wall consisted of a response panel with three keys situated 21cm above the chamber floor. The center key was located 16cm from each wall and the other two keys 8cm on either side of the center key. Only the center and right keys were used during sessions. The center key was lit red and the right key green, each key required a force of approximately 0.15N for a response to be registered. The panel also contained a house-light situated 8cm above the center key, and a food hopper, situated 13cm below it. During reinforcement the houselight and response keys were dark, and the hopper was illuminated and raised to allow access to the wheat. During experimental sessions fans, attached to each chamber, provided ventilation and masking noise. Sessions were controlled and recorded by a computer running MED-PC software, which was situated in an adjacent room.

Procedure

Baseline Training:

Sessions were run at approximately the same time of day, seven days per week. Throughout training sessions consisted of 105 trials, with 35 trials of each of three different types (2, 4 and 6 flashes). Trials were determined pseudorandomly, with the constraint that out of every nine trials, there were three of each type. Each trial consisted of two phases; a sample phase and a production phase. During the sample phase a number of flashes (2, 4 or 6) were presented to the subject, and they had to re-produce that number during the production phase.

Sample phase:

There were two versions of the sample phase termed 'time' and 'rate' controlled. Birds were exposed to either rate-controlled or time-controlled baseline training in different conditions. In the time-controlled procedure, all three flash sequences were presented within a ten-second interval i.e., 2-flashes/10-seconds, 4-flashes/10-seconds and 6-flashes/10-seconds. This procedure was designed to minimise temporal cues and promote subjects' attention to numerical cues. In the rate-controlled procedure, one flash occurred every 2.5 seconds, so that 2-flashes took 5 seconds, 4-flashes 10 seconds and 6-flashes 15 seconds. In this procedure, time and number were both equally good predictors of the response requirement in the production phase. An average response latency (to peck the sample key) of 1-second was assumed when the flash intervals were programmed. Thus, for example, on a two-flash trial in the time-controlled procedure after four seconds the centre key would be lit red, and after nine seconds it would be lit again.

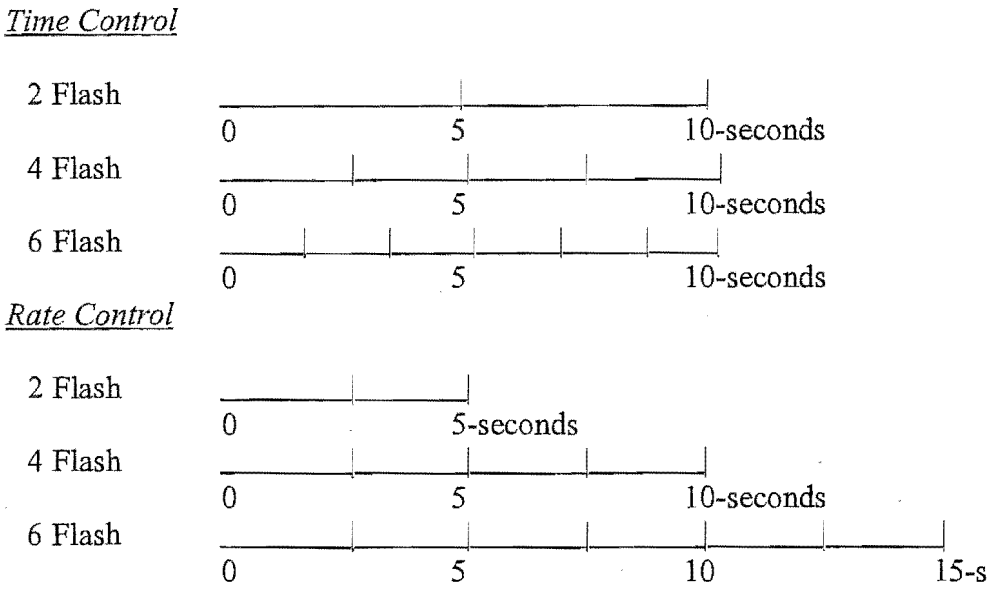

Figure 4. This figure illustrates the presentation of flashes in the rate and time control procedures. In the time control procedure all flashes are presented in a 10-second interval, while in the rate control procedure one flash is presented every 2.5-seconds.

At the beginning of the flash phase, the house light was illuminated, and after a fixed duration (depending on the flash sequence) the centre key was lit red. One peck to the key turned the key light off until the next flash was scheduled, when it was again illuminated. This continued until the designated number of flashes for that sequence had been completed. Following the final flash, there was a 2-second retention interval during which all key lights and the house light were turned off.

Production phase:

After the retention interval, the house light was turned on and the centre and right keys were lighted red and green respectively. In order to obtain reinforcement, subjects had to produce the same number of pecks as presented in the flash phase on the centre key, and then peck the right key once to signal that they had completed the trial. Thus, if four flashes were presented on the centre key, following the retention interval, subjects had to peck the centre key four times before switching to the right

key and pecking once. When the right key had been pecked, the trial ended and both keylights were turned off.

If subjects made the correct number of centre key pecks before switching to the right key, the hopper light came on and the hopper was raised for 4.5-seconds. This was followed by a 12-second inter-trial interval (ITI), during which the houselight and keylights were turned off. After the ITI the next trial began.

If subjects made an incorrect number of centre key responses followed by a right-key response, no reinforcement was delivered. Instead, a 5-second ITI began, after which subjects were given a correction trial. Correction trials were identical to the preceding regular trial, except during the production phase when only the centre key was lit red. Once the correct number of centre key pecks had been made, the centre key was turned off and the right key was illuminated green. One peck to the right key turned the key off and 1.5-seconds of reinforcement was delivered. Following a 12-second ITI the next trial began. All dark-key responses produced a 1.5-second delay.

Testing:

Following baseline training, testing was conducted with novel probe trials. These consisted of 1, 3, 5 and 7 flashes presented in the manner described in baseline training. Each session of testing consisted of 85 baseline trials interspersed with 20 probe trials (5 of each type), with the identity of each trial determined pseudorandomly. Reinforcement was delivered randomly on these probe trials; the probability of reinforcement was the same as the average obtained probability of reinforcement on regular trials over the preceding five baseline sessions (see table 1).

As in training, reinforcement on baseline trials was delivered following a correct response, with correction trials following incorrect responses.

Two types of testing were carried out: consistent transfer tests and inconsistent transfer tests. In consistent transfer tests, baseline (2, 4, and 6 flashes) and probe (1, 3, 5 and 7 flashes) trials were presented in the same manner. Thus, if subjects had just completed time-controlled baseline training, both baseline and probe trials were presented in a time-controlled manner, and vice versa. In inconsistent transfer tests, subjects received probe trials in the opposite manner to their preceding baseline training. For example, baseline trials would be time-controlled while probe trials would be rate-controlled, and vice versa.

To the extent that subjects have acquired numerical competence a high degree of transfer should be obtained in these tests. While the time-controlled procedure was designed to minimise temporal information, baseline conditions are not sufficient to rule out timing. Subjects may rely on temporal cues other than total sample duration, such as inter-stimulus intervals. A critical question is the extent to which probe responding is disrupted when subjects are given inconsistent tests. Performance on these trails may indicate whether subjects' behaviour was determined by numerical or temporal cues.

Table 1. Reinforcement probability on probe trials.

	Q5	Q6	Q7	Q8
Condition 1	0.273	0.312	0.278	0.329
Condition 2	0.225	0.232	0.265	0.167

Condition 1:

In Condition 1, birds Q5 and Q6 were trained under the time-controlled procedure, while Q7 and Q8 were trained under the rate-controlled procedure. All birds, except Q5, were exposed to the full procedure (i.e. all three flash sequences) at the beginning of training. Bird Q5 was initially given only two-flash trials. Six-flash trials were then introduced and when Bird Q5 showed differentiation between two and six, four-flash trials were added. During preliminary training, in some sessions, dark key responses produced timeouts (1.5-seconds), but because this was found to disrupt performance the timeouts were removed. Numbers of sessions in baseline training for each subject are shown in Table 2. Some subjects began to show stereotyped responding, making two responses on all three flash sequences, so the proportions of 2, 4 and 6 trials were altered in an effort to encourage subjects to discriminate between them (see Table 3).

After baseline training, the consistent transfer test was conducted. For birds Q5 and Q6, who received time-controlled training, all trials during testing (baseline and probe) were presented in a time-controlled manner. Conversely, Q7 and Q8 received both types of trials in a rate-controlled manner. This was followed by a return to baseline conditions (2, 4 and 6-flash sequences only) for ten sessions. Following this subjects were given inconsistent transfer tests, in which Q5 and Q6 received time-controlled baseline trials (2, 4 and 6 flashes) but rate-controlled probe trials (1, 3, 5, and 7 flashes), while Q7 and Q8 received rate-controlled baseline trials and time-controlled probe trials. Because of a programming error, Q8 received consistent, time-controlled, rather than rate-controlled trials in the initial transfer test. He was returned to baseline for ten sessions and then received the appropriate tests.

Condition Two:

After testing in Condition 1, subjects received baseline training in the opposite procedure. Thus, Q5 and Q6 received rate-controlled training and Q7 and Q8 received time-controlled baseline training. Q7 and Q8 initially received only two-flash trials, with six- and four-flash trials successively added. Training was again followed by consistent transfer tests, a return to baseline, and then inconsistent transfer tests (see table 2).

Table 2. Number of sessions during training and testing for Q5-8.

Condition 1	**Q5**	**Q6**	**Q7**	**Q8**
Baseline	111	101	101	111
Consistent transfer	10	10	10	10
Baseline	10	10	10	10
Inconsistent transfer	10	10	10	10
Condition 2				
Baseline	51	65	61	41
Consistent transfer	10	10	10	10
Baseline	11	11	11	10
Inconsistent transfer	10	10	10	10

Table 3. Proportions of type 2, 4 and 6 trials received during the final ten sessions of training and during testing.

	Q5			Q6			Q7			Q8		
	2	4	6	2	4	6	2	4	6	2	4	6
Condition 1												
Baseline	0.44	0.11	0.44	0.11	0.44	0.44	0.11	0.44	0.44	0.33	0.33	0.33
Consistent	0.45	0.10	0.45	0.10	0.45	0.45	0.10	0.45	0.45	0.33	0.33	0.33
Baseline	0.44	0.11	0.44	0.11	0.44	0.44	0.11	0.44	0.44	0.11	0.44	0.44
Inconsistent	0.45	0.10	0.45	0.10	0.45	0.45	0.10	0.45	0.45	0.10	0.45	0.45
Condition 2												
Baseline	0.44	0.22	0.33	0.22	0.33	0.44	0.22	0.33	0.44	0.22	0.33	0.44
Consistent	0.45	0.21	0.34	0.21	0.34	0.45	0.21	0.34	0.45	0.21	0.34	0.45
Baseline	0.44	0.22	0.33	0.22	0.33	0.44	0.22	0.34	0.44	0.22	0.33	0.44
Inconsistent	0.45	0.21	0.34	0.21	0.34	0.45	0.21	0.34	0.45	0.21	0.34	0.4

RESULTS

The primary data analysed were the number of center key pecks made during the production phase. Initially, percent correct in the two procedures was analysed. Then, response distributions were examined from a qualitative perspective. Finally, the quantitative indices of discrimination performance seen in the response distributions were examined.

The percentage of correct responses made following each flash sequence was determined using data, averaged over conditions, from the last five sessions of baseline training in both the rate and time-controlled procedures. These results are illustrated in Figure 5.

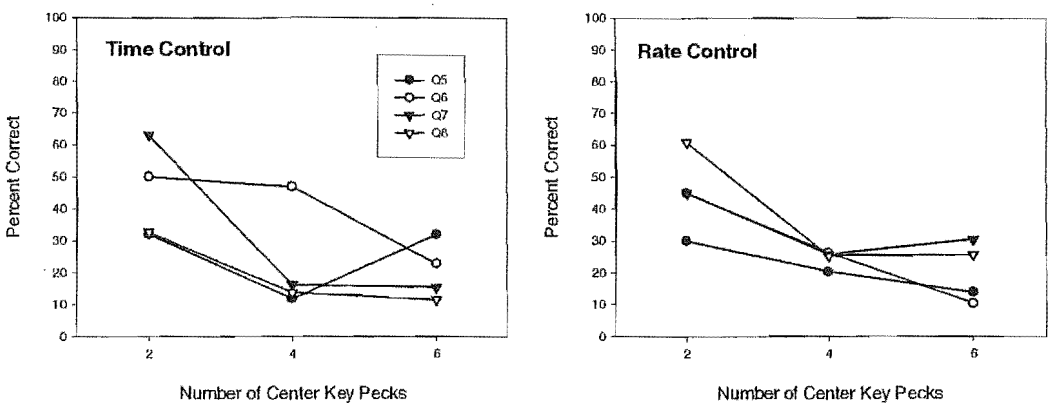

Figure 5. Percent correct as a function of number of center key pecks made during the production phase in time (left graph) and rate (right graph) controlled procedures from the last five sessions of baseline training for subjects Q5-Q8.

Performance in the three trial types ranged from 10 to 60 percent correct across both conditions. All subjects made the fewest errors on two-flash trials and, with the exception of Q5 in the time controlled procedure and Q8 in the rate-controlled procedure, made more correct responses on four-flash trials than six-flash trials. Thus, accuracy, in terms of percent correct, generally decreased as the number of flashes in the sample phase increased. A 2x3 repeated-measures ANOVA was conducted on the percent correct data with condition (rate vs. time) and number of stimuli (2, 4, 6) as factors. There was a significant effect of number of stimuli, $F(2,6) = 11.73$, $p < .01$. The main effect of condition, and the interaction between condition and number of stimuli failed to reach significance. A second repeated-measures ANOVA was conducted to examine possible differences between Conditions 1 and 2. The main effect of condition (Condition 1 vs. Condition 2), and the interaction between condition and number of stimuli also failed to reach significance. A planned linear contrast on number of stimuli revealed a significant effect, $F(1,3) = 19.75$, $p < .03$. This demonstrates that percent correct decreased linearly as the number of stimuli increased.

Percent correct, however, is a relatively crude measure of accuracy. Of interest is how many center key responses subjects' made in the production phase following each of the three flash sequences. Response distributions were produced by determining the proportion of trials on which subjects made zero through twelve responses for each flash sequence.

Baseline Trial Performance

This section concentrates on baseline trial performance (2, 4 and 6-flash sequences) during training and testing is examined. Probe trial performance (1, 3, 5 and 7-flash sequences) is examined in the next section. In the first condition, subjects Q5 and Q6 responded in the time-controlled procedure, while Q7 and Q8 responded in the rate-controlled procedure. The left-hand column in Figure 6 shows response distributions for two, four, and six-flash trials from the last five sessions of baseline training in Condition 1. Filled circles represent two-flash trials, open circles represent four-flash trials and filled triangles represent six-flash trials. All four subjects show three reasonably distinct response distributions with peaks at, or near, the reinforced number for that trial. Two-flash trial performance was the most accurate, with all subjects showing peaks at two responses. In terms of percent correct (Figure 5), Q6 was more accurate on four-flash trials than on six-flash trials, while Q8 was equally accurate on four and six-flash trials. Figure 6 shows that both Q6 and Q8 had peaks at five responses on six-flash trials. The opposite was true for Q5 and Q7, who were less accurate on four-flash trials and showed peaks at six responses on six trials.

After baseline training, transfer testing with consistent probe trials was conducted. The left-hand column of Figure 7 shows response distributions for baseline trials (i.e. two- four- and six-flash sequences) during the last five sessions of consistent transfer testing. The introduction of novel probe trials disrupted baseline trial performance, with the general trend being a decrease in accuracy. With the exception of Q6, all subjects were less accurate on two-flash trials during testing. On four-flash trials, Q7 and Q8's performance improved, with response distribution peaks moving from five to four responses. Q5's four-flash trial performance remained stable while

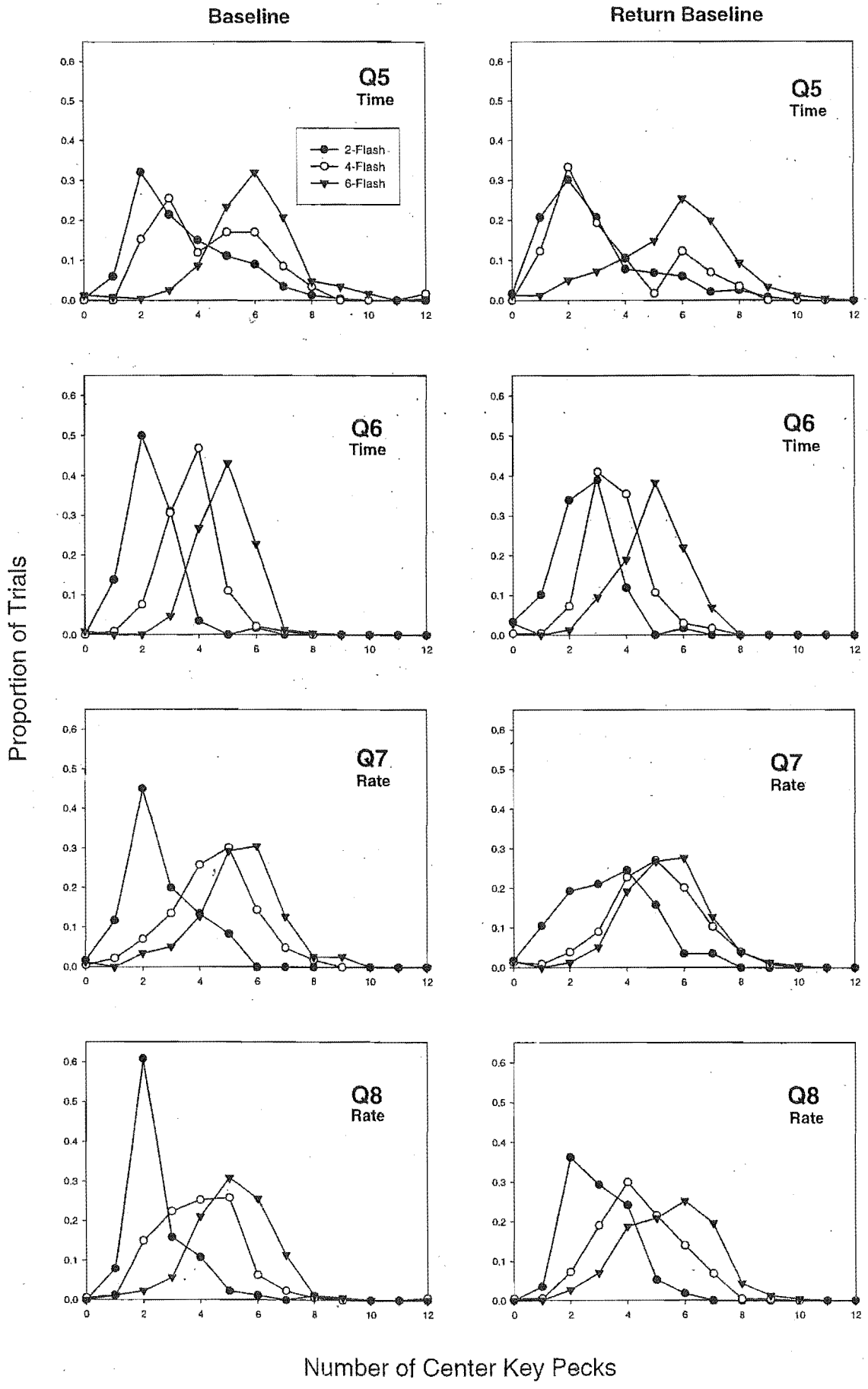

Figure 6. Response distributions for 2, 4 and 6-flash trials from the last five sessions of baseline training and return to baseline (following consistent transfer tests) in Condition 1. Q5 and Q6 responded in the time-controlled procedure while Q7 and Q8 responded in the rate-controlled procedure.

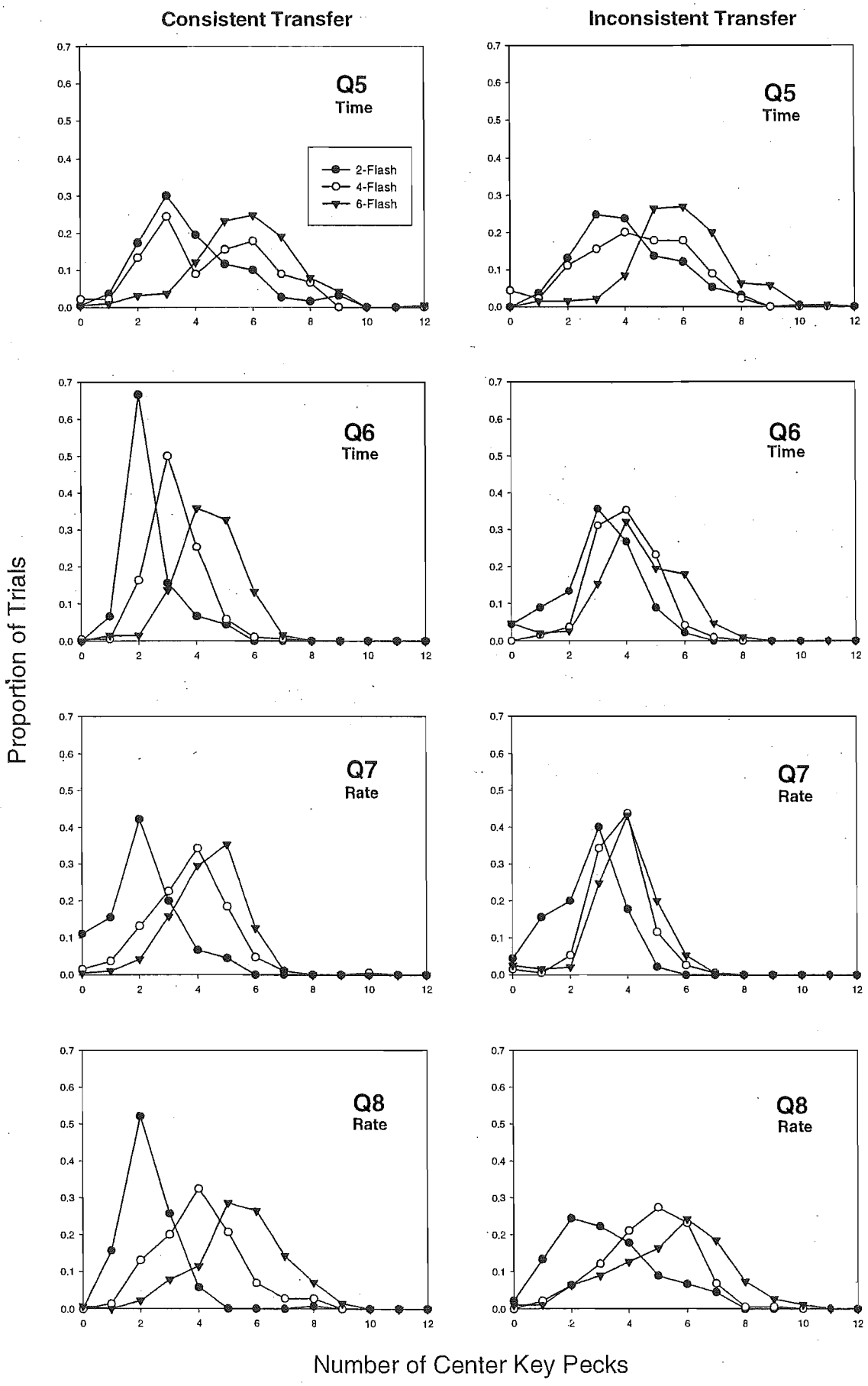

Figure 7. Response distributions for 2, 4 and 6-flash trials from the last five sessions of consistent (left column) and inconsistent (right column) transfer testing in Condition 1.

Q6's performance worsened. Only Q8 maintained its previous six-flash trial performance during testing.

Returning to Figure 6, the right-hand column shows response distributions for two, four, and six-flash trials when subjects were returned to baseline for ten sessions following consistent transfer testing. Subjects' performance generally recovered from the disruption produced by the addition of the novel probe trials, except for two-flash trial performance for Q6, Q7, and to a lesser extent, Q8.

Following the return to baseline, inconsistent transfer tests were conducted. The right hand column in Figure 7 shows response distributions for two, four, and six-flash trials in the last five sessions of inconsistent transfer testing. The addition of the inconsistent probe trials appeared to disrupt baseline trial performance to a greater extent than did the addition of consistent probe trials. For birds Q6 and Q7, much discrimination between two, four, and six-flash trials was lost, with the response distributions overlapping substantially. Q5 lost accuracy on two-flash trials and Q8 on four-flash trials.

In the second condition, subjects were given training in the opposite procedure to that in Condition 1. Thus, Q5 and Q6 received rate-controlled baseline training and Q7 and Q8 received time-controlled baseline training. The left-hand column of Figure 8 shows response distributions for baseline trials from the final five sessions of training in Condition 2. As in Condition 1, all subjects' showed a peak at two responses on two-flash trials. However, overall subjects' performance appeared to be worse than in Condition 1 with less differentiation between the three flash sequences. This may have been due to the smaller number of training sessions given in Condition 2. This is supported by the fact that Q8 who had the fewest number of training

sessions also had the poorest performance, both compared to its own Condition 1 performance and also to that of the other three subjects in Condition 2.

The left-hand column of Figure 9 shows response distributions for baseline trials from the last five sessions of consistent transfer testing in Condition 2. During the consistent transfer test, baseline performance improved for Q6, in that distributions for the baseline trials separated from one another. For Q7 and Q8 the addition of the probe trials disrupted performance slightly. Both birds lost some of the discrimination on two-flash trials and their response distributions overlapped to a greater extent.

Following the consistent transfer tests, subjects were again returned to baseline conditions for ten sessions, and data from the last five are illustrated in the right-hand column of Figure 8. For three of the four subjects (Q5, Q6 and Q7) baseline performance following the transfer test improved, although performance was still less accurate than it was in Condition 1. This is consistent with the idea that the smaller number of sessions given in Condition 2 training was responsible for subjects' poorer performance.

Figure 9 (right column) shows response distributions for baseline trials during the last five sessions of inconsistent transfer testing. In contrast to the findings of Condition 1, baseline performance showed little disruption compared with preceding baseline training. Q8, whose baseline performance in Condition 2 was relatively poor, showed a dramatic improvement on baseline trials during inconsistent testing. The three response distributions separated from one another and showed peaks at the reinforced number of responses. The reasons for this improvement are unclear.

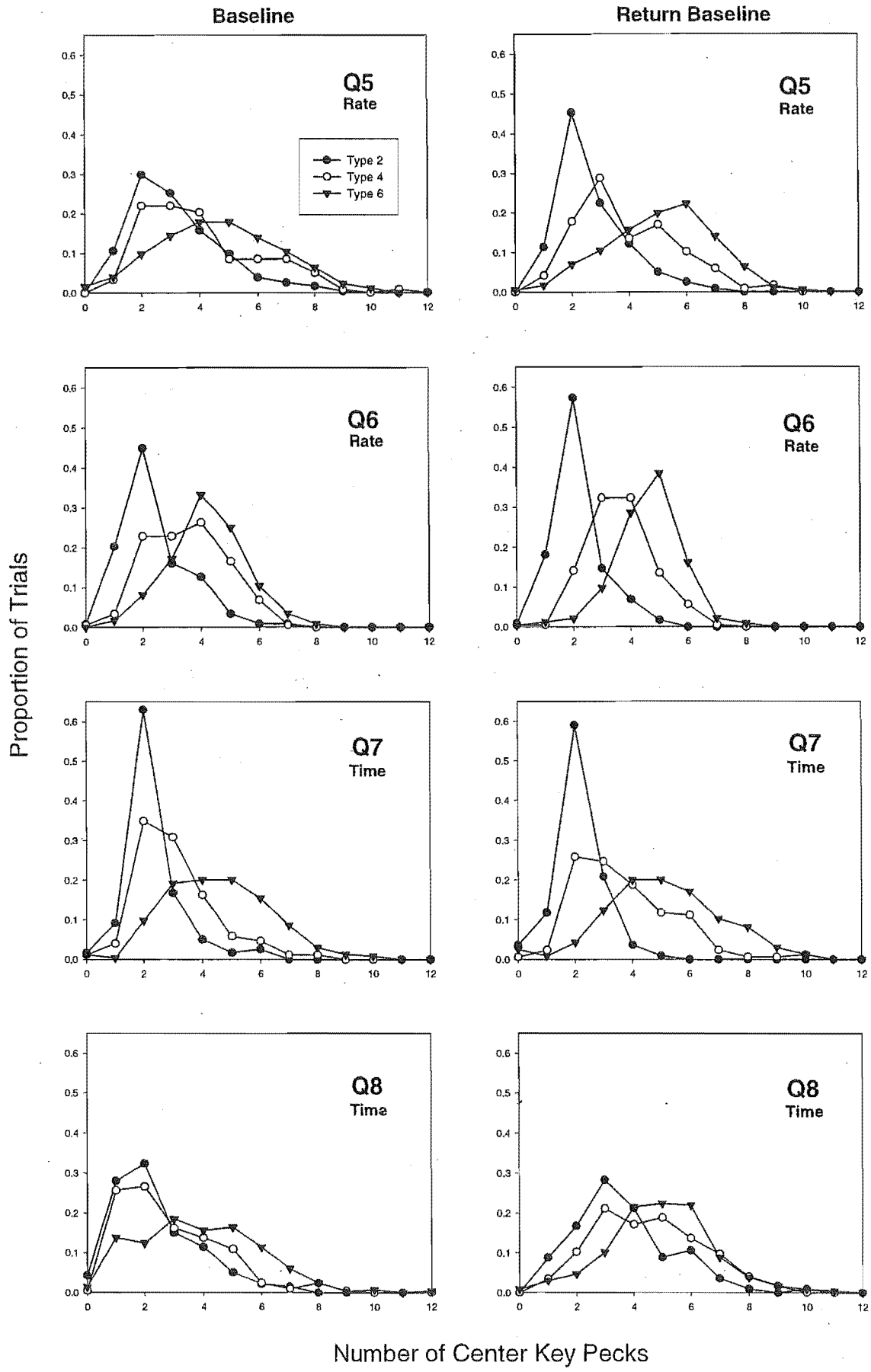

Figure 8. Response distributions for 2, 4, and 6-flash trials from the last five sessions of baseline training and return to baseline (following consistent transfer tests) in Condition 2. Q5 and Q6 responded in the rate-controlled procedure while Q7 and Q8 responded in the time-controlled procedure.

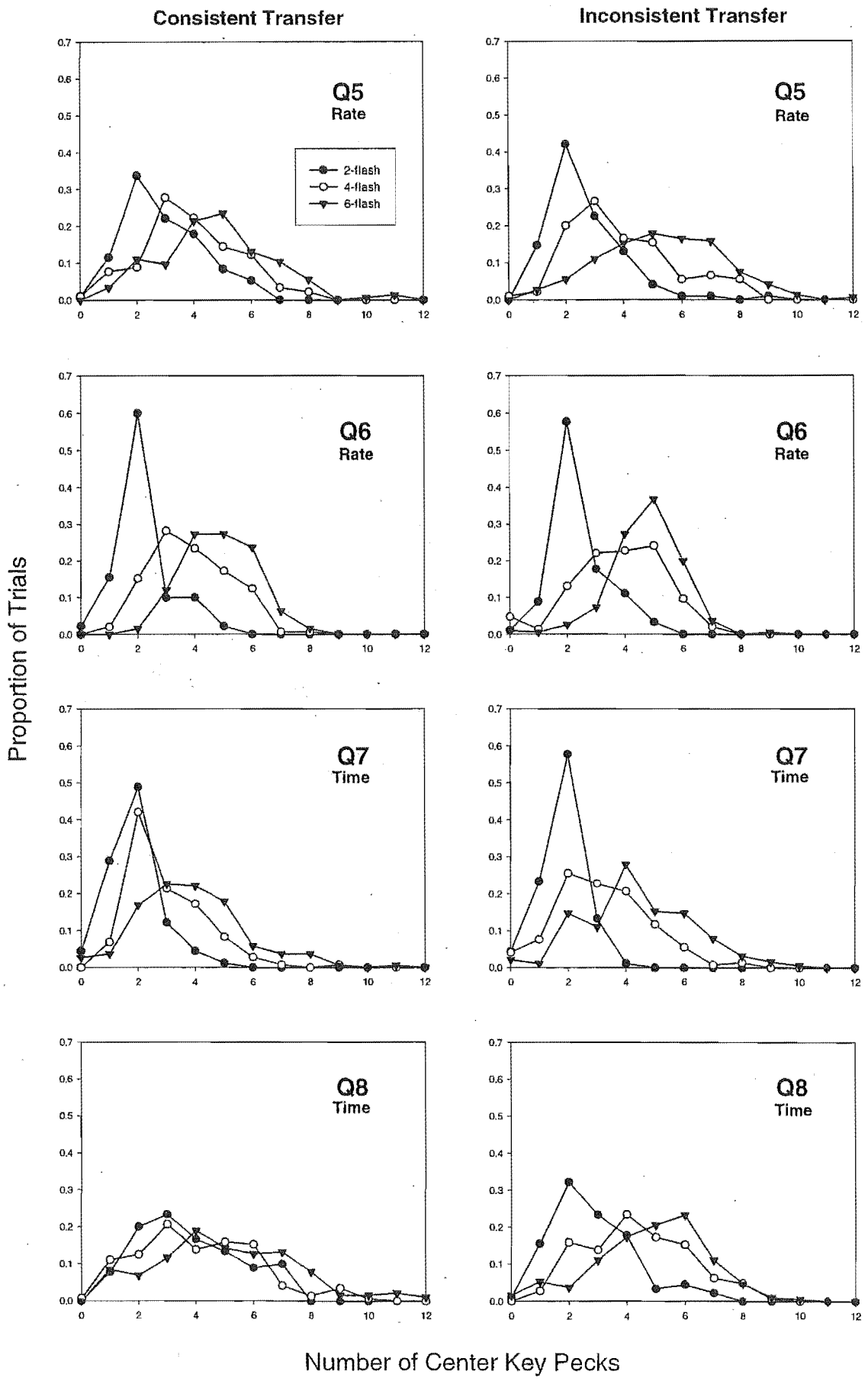

Figure 9. Response distributions for 2, 4 and 6-flash trials from the last five sessions of consistent (left column) and inconsistent (right column) transfer testing in Condition 2.

One way to quantify the disruptive effects of novel probe trials on baseline performance is to compare the slopes of functions plotting average number of responses against number of flashes for baseline trials during training and testing. Perfect performance would imply a slope of 1.0. Actual slopes not only illustrate the overall accuracy of discrimination between the trials but also demonstrate the degree to which baseline performance was disrupted by the addition of probe trials: as the disruption to baseline performance increases, slope values decrease. Table 4 gives slope values for baseline trials during initial baseline training, consistent transfer tests, return to baseline and inconsistent transfer tests for all subjects.

Table 4. Slope values for baseline training (BL), return to baseline (RBL), consistent transfer test (CT) and inconsistent transfer tests (ICT) in Conditions 1 and 2.

		Condition 1	Condition 2
Q5	BL	0.65	0.58
	RBL	0.61	0.57
	CT	0.60	0.49
	ICT	0.56	0.62
Q6	BL	0.58	0.46
	RBL	0.57	0.66
	CT	0.53	0.62
	ICT	0.30	0.58
Q7	BL	0.72	0.51
	RBL	0.57	0.58
	CT	0.51	0.49
	ICT	0.31	0.61
Q8	BL	0.56	0.28
	RBL	0.60	0.30
	CT	0.83	0.26
	ICT	0.60	0.47

The slope values shown in Table 4 are plotted in Figure 10. In Condition 1, baseline trial performance in the consistent transfer test was disrupted by the addition of probe trials for three of four subjects. All subjects recovered some, but not all, of their

performance when returned to baseline. For Q6 and Q7 the inconsistent probe trials disrupted baseline performance during the test to a greater extent than the consistent probe trials. Slopes from Condition 2 similarly indicate that performance was disrupted by the addition of consistent probe trials for subjects Q5 and Q7. However, baseline performance improved following the consistent transfer test for these subjects (relative to their baseline trial performance during the test). Also in contrast to Condition 1, for three of the four subjects, the inconsistent probe trials produced less disruption than the consistent probe trials, particularly for Q8.

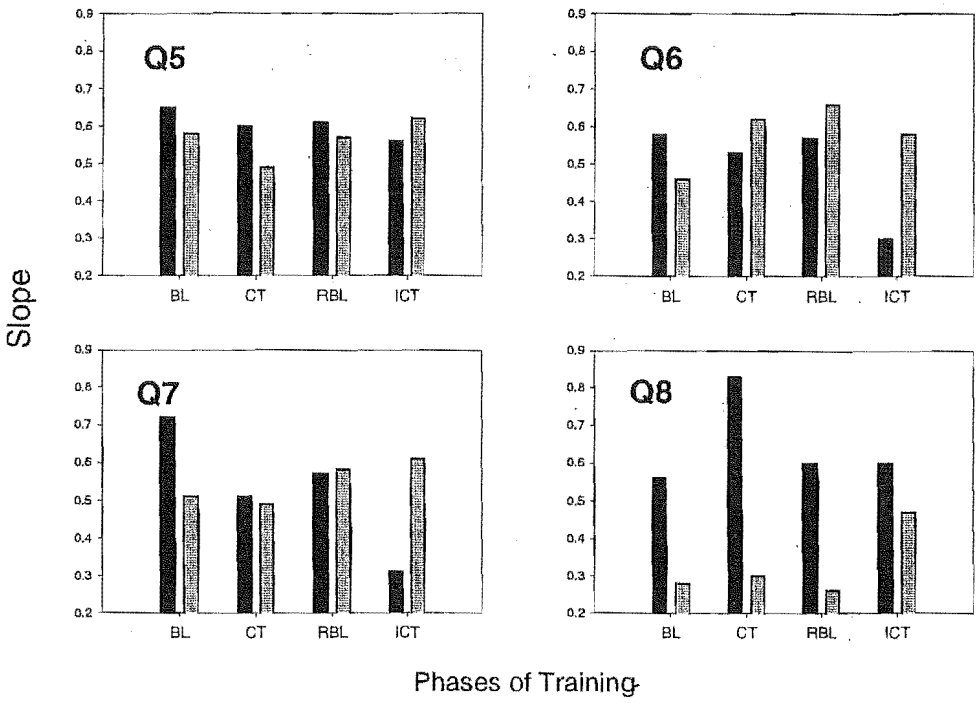

Figure 10. Slope values are plotted for baseline trials during training (BL), consistent transfer testing (CT), return to baseline (RBL) and inconsistent transfer testing (ICT). Black bars represent Condition 1 and gray bars represent Condition 2.

Probe Trial Performance

Response distributions for the novel 1, 3, 5 and 7-flash probe trials in consistent transfer testing in Condition 1 are shown in Figure 11 and the left-hand column of Figure 12. The left-hand column of Figure 11 shows data from the first five sessions of testing and the right-hand column shows data from the last five sessions of testing. Figure 12 shows data from all ten sessions of testing. Filled circles and triangles represent 1 and 3-flash trials respectively, while unfilled circles and triangles represent 5 and 7-flash trials, respectively. Figure 11 shows that performance on the probe trials generally improved over the ten sessions, with the exception of Q5 who failed to develop clear differentiation among the four flash sequences. For the other three subjects, response distributions for 1 and 3-flash trials (filled symbols) were clearly different from those for 5 and 7-flash trials (open symbols). Overall subjects were most accurate on 3- and 5-flash trials and least accurate on 7-flash trials. Comparison of probe trial performance in Figure 12 (left column) and baseline performance in Figure 7 (left hand column) suggests that to some extent subjects were discriminating between baseline and probe trials, despite response distribution peaks for the probe trials not lying at the target number of center key pecks. All subjects made one response on 1-flash trials more than they did on 2-flash trials, indicating differentiation between the two sequences. Q7 and Q8 made three responses on 3-flash trials more than on either 2- or 4-flash trials, while Q6 made three responses on 3-flash trials more than on 2- but not 4-flash trials. Q5 and Q7 made five responses on 5-flash trials more than they did on 4- but not 6-flash trials. In fact, Q7 showed peaks at 1, 2, 3, 4, and 5 responses on 1, 2, 3, 4 and 5-flash trials respectively (see Fig. 7 and Fig. 11). Q8's response

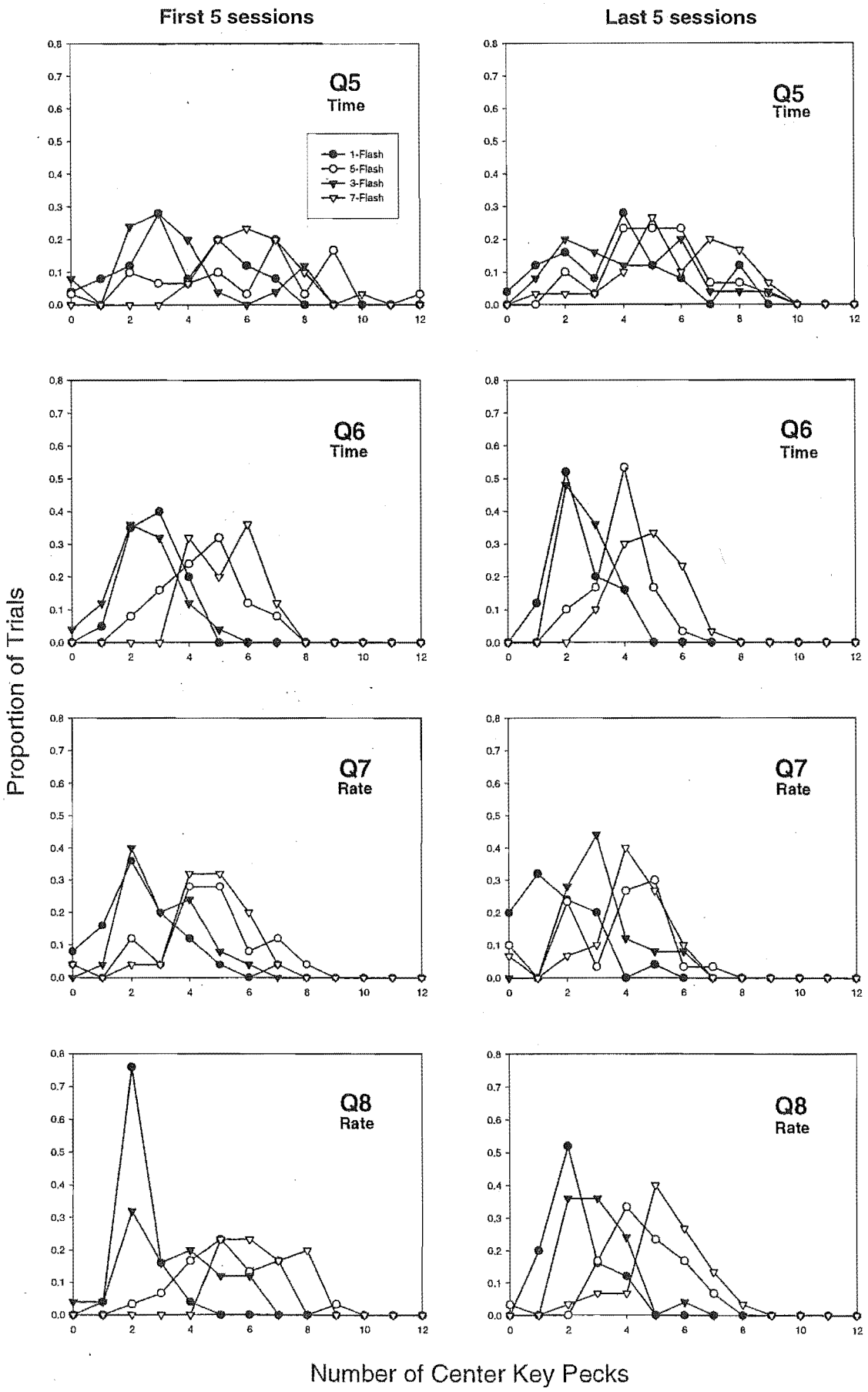

Figure 11. Response distributions for 1, 3, 5 and 7-flash trials from the first five (left column) and last five (right column) sessions of consistent transfer testing in Condition 1.

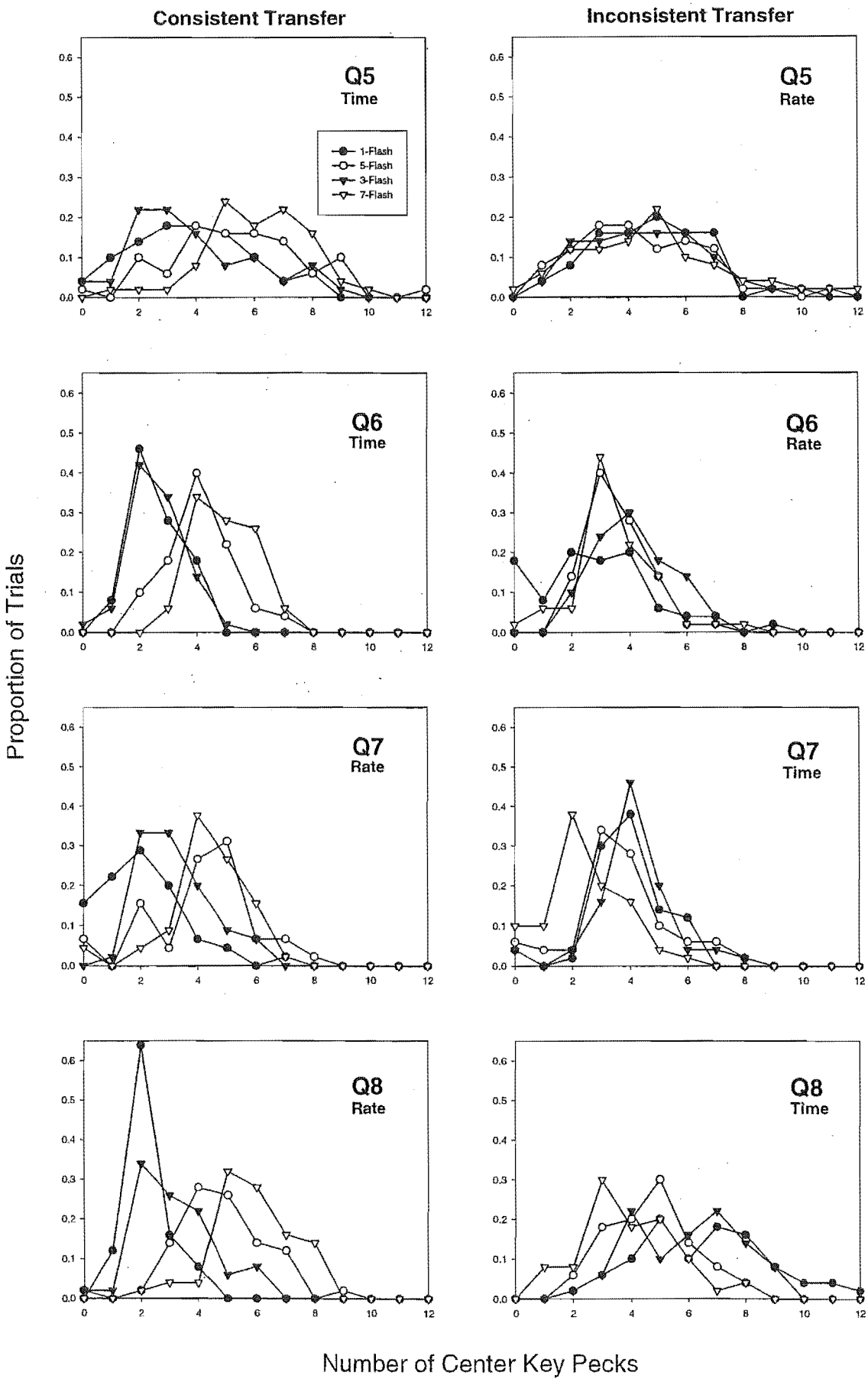

Figure 12. Response distributions for 1, 3, 5 and 7-flash trials from all 10 sessions of consistent (left column) and inconsistent (right column) transfer testing in Condition 1.

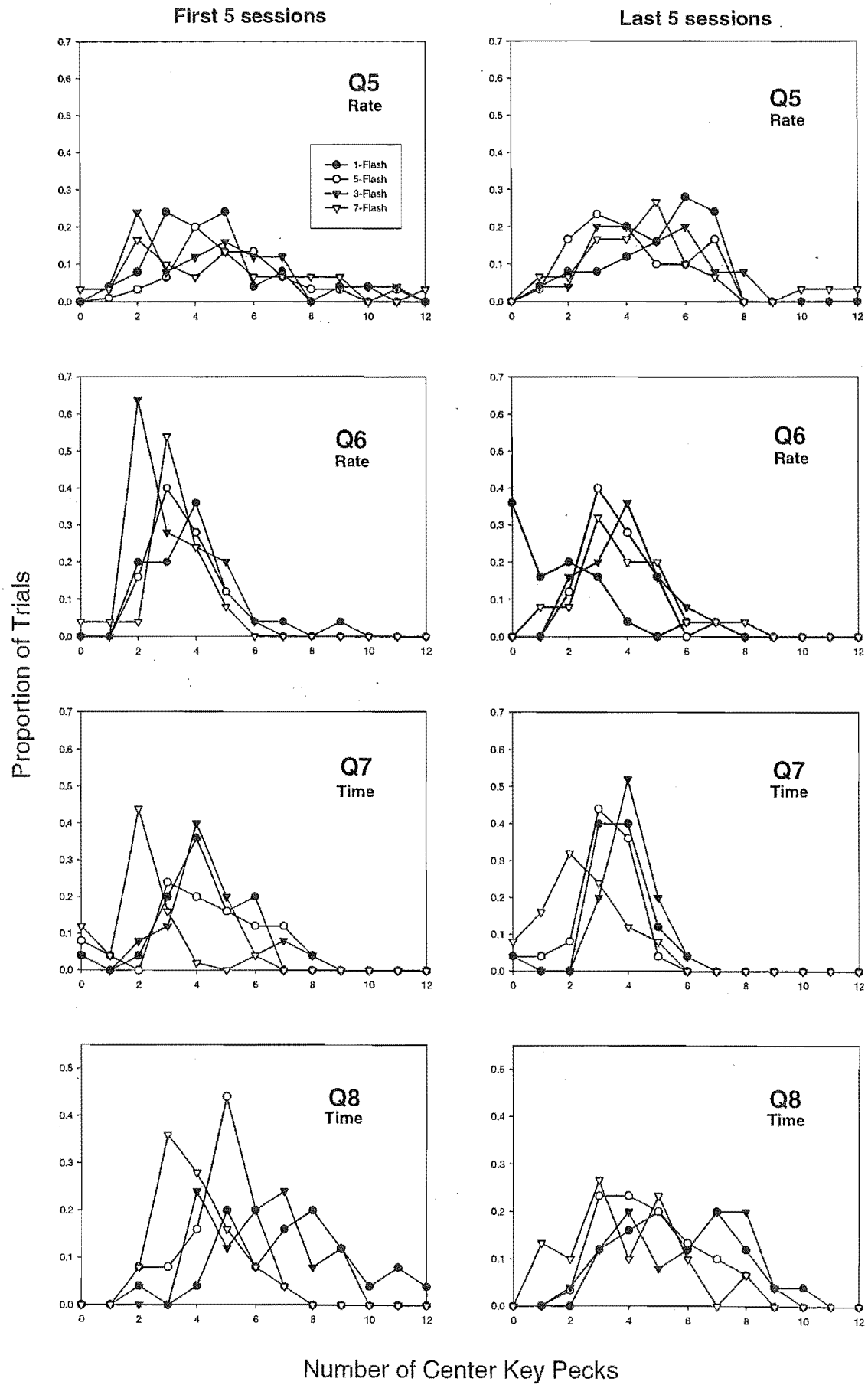

Figure 13. Response distributions for 1, 3, 5 and 7-flash trials from the first five (left column) and last five (right column) sessions of inconsistent transfer testing in Condition 1.

distribution for 5-flash trials was almost identical to its 2-flash distribution. On 7-flash trials, only Q8 made seven responses more than on 6-flash trials.

Inconsistent testing involved probe trials being presented in the opposite procedure to that of the baseline trials. In Condition 1, Q5 and Q6 received time-controlled baseline trials and rate-controlled probe trials while Q7 and Q8 received rate-controlled baseline trials and time-controlled probe trials. Figure 13 and the right hand column of Figure 12 show response distributions for probe trials during inconsistent transfer testing in Condition 1. Performance did not change consistently over the ten sessions of testing with little difference in the graphs illustrating the first five, last five or all ten sessions of data. Performance on the inconsistent probe trials was also poorer than that on the consistent probe trials and response distributions overlapped to a greater extent. This was particularly true for Q6 and Q8, while Q5 showed little discrimination between probe trials on either test.

In Condition 2, the training regimes were reversed for all subjects, and consistent and inconsistent tests were repeated after a second baseline was established. Figure 14 shows response distributions for probe trials from the first five sessions (left column) and the last five sessions (right column) of consistent transfer testing in Condition 2. Overall, performance on all four flash sequences improved over the ten sessions of testing, 1 and 3-flash distributions separated from 5 and 7-flash distributions, with the exception of 3- and 7-flash performance for Q6. Figure 15 (left-hand column) shows response distributions for 1, 3, 5 and 7-flash trials from all ten sessions of testing, and best represents subjects' performances. Q8 shows little discrimination between baseline (see Figure 9) or probe trials during testing, with all response distributions lying close together. Although the four response distributions

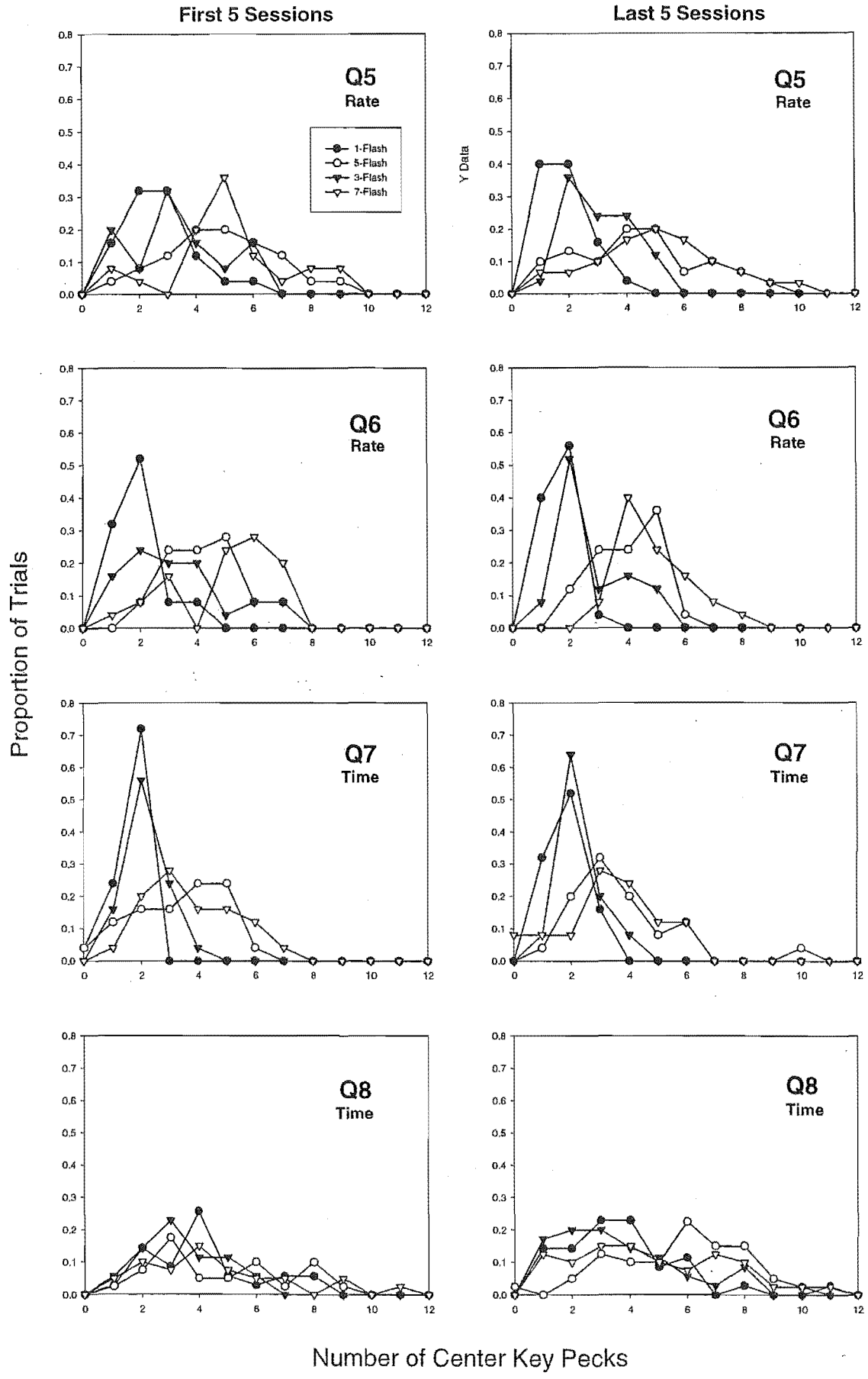

Figure 14. Response distributions for **1**, 3, 5 and 7-flash trials from the first five (left column) and last five (right column) sessions of consistent transfer testing in Condition 2.

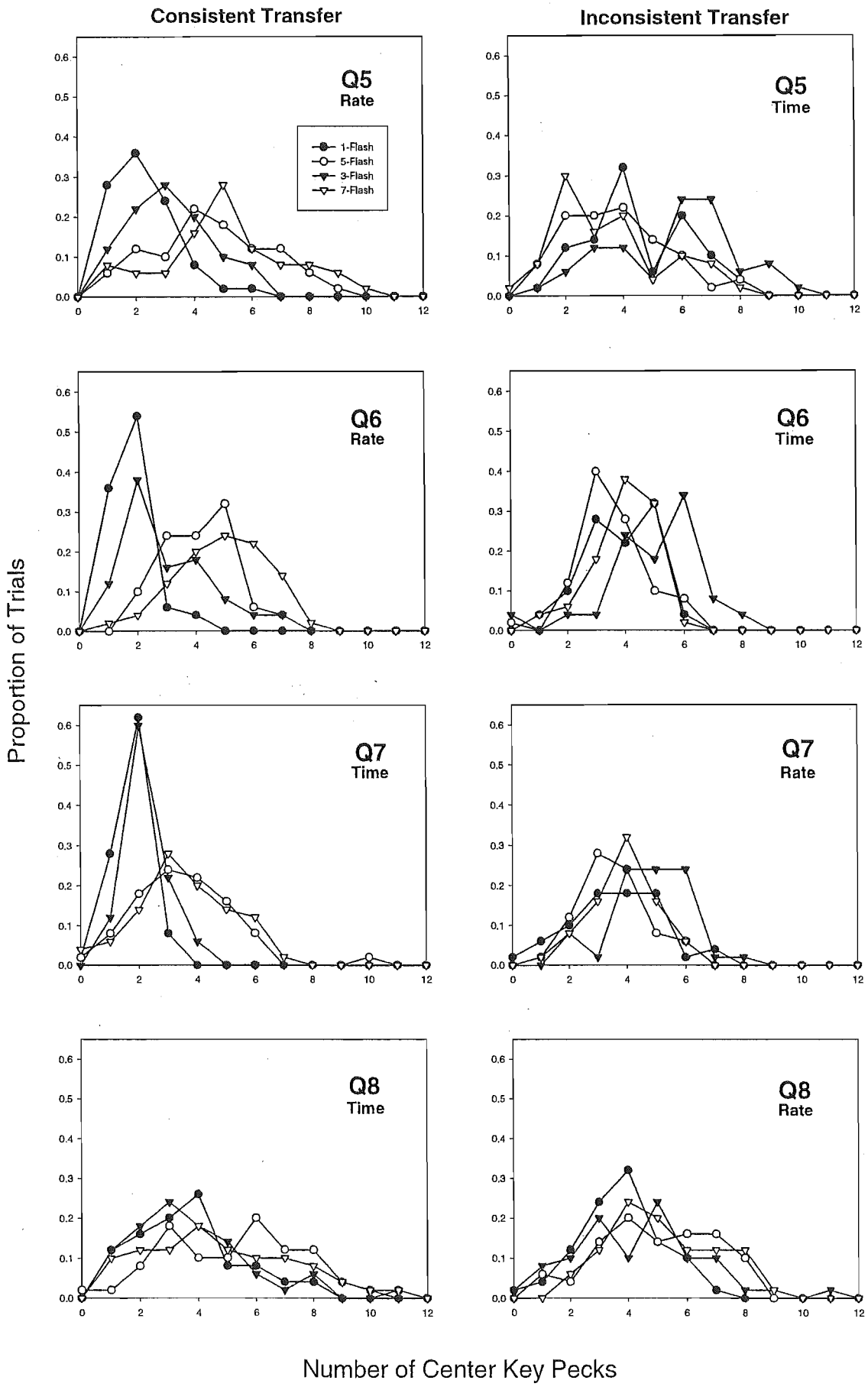

Figure 15. Response distributions for 1, 3, 5 and 7-flash trials from all 10 sessions of consistent (left column) and inconsistent (right column) transfer trials in Condition 2.

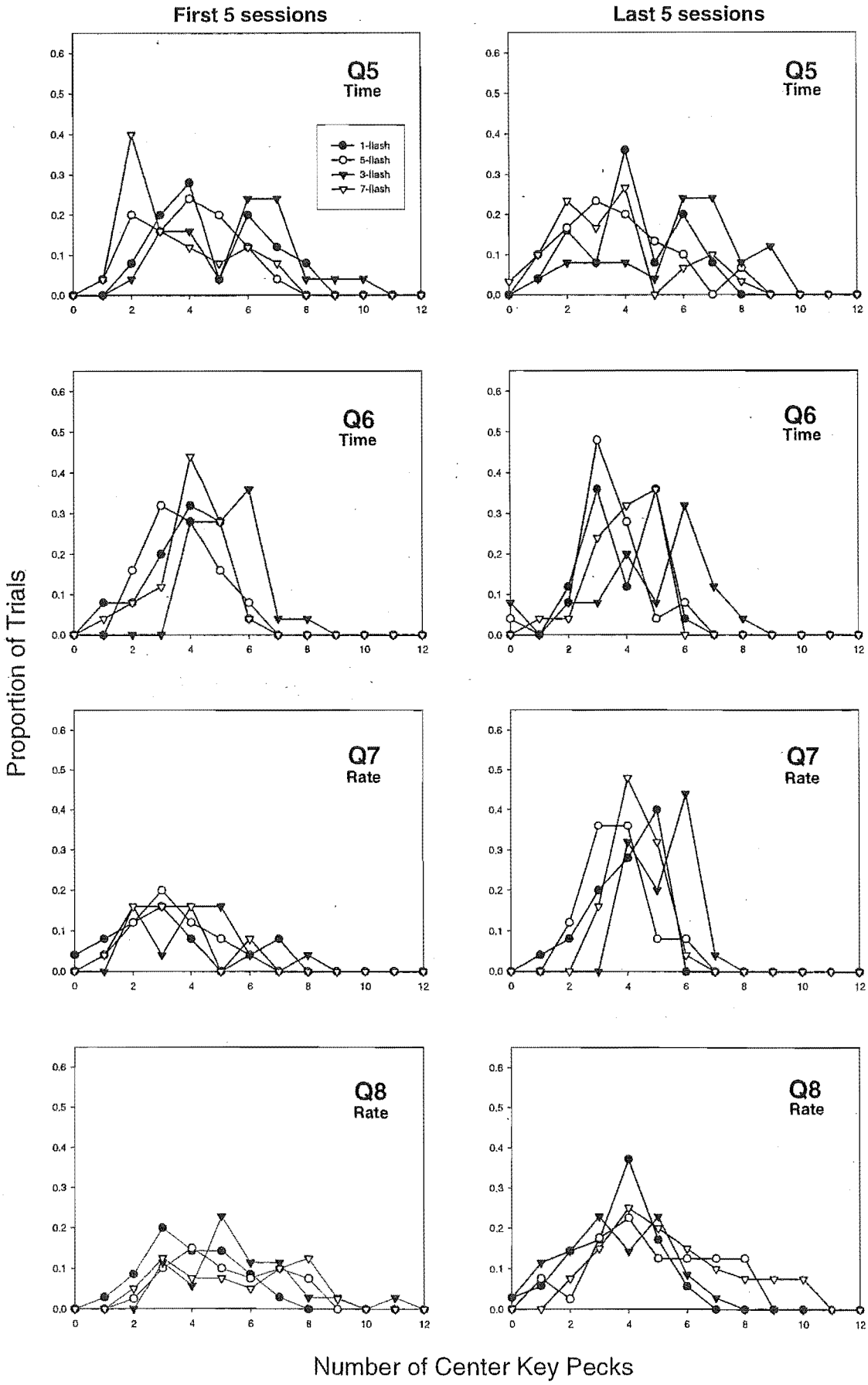

Figure 16. Response distributions for 1, 3, 5 and 7-flash trials from the first five (left column) and last five (right column) sessions of inconsistent transfer testing in Condition 2.

(Figure 15) for Q5 and Q6 generally did not peak at the target number of responses again there is some indication that subjects differentiated between the baseline and probe trials. Both Q5 and Q6 show some discrimination between 1 and 2-flash trials, making more one-peck responses on the former than the latter. Q6 showed some discrimination between 4, 5 and 6-flash trials, making five responses more frequently on 5-flash trials than either 4 or 6-flash trials. Q6 also showed discrimination between 6 and 7-flash trials, making seven responses almost three times as often on 7-flash trials as on 6-flash trials. Q7 appeared to treat 1 and 3-flash trials as equivalent to 2-flash trials, and 5 and 7-flash trials as equivalent to 6-flash trials with response distributions for the probe trials almost identical to the respective baseline distributions. However, 1 and 3-flash trial distributions for Q7 are different from those for 5 and 7-flash trials indicating discrimination between large and small numbers of flashes.

Figure 16 shows response distributions for 1, 3, 5 and 7-flash trials in the first five (left column) and last five (right column) sessions of inconsistent transfer testing. As in Condition 1, response distributions for inconsistent probe trials overlapped to a greater extent than those for consistent probe trials. Performance did not significantly change over the ten sessions, except for Q7 whose response distributions for 5 and 7-flash trials (open symbols) separated from 1 and 3-flash trial distributions (filled symbols) over sessions of testing.

Clearly, several subjects discriminated between the four probe trials to some extent. Figures 17 and 18 show the number of responses (averaged over all 10 sessions) made on baseline and probe trials in both consistent and inconsistent transfer tests in Condition 1 and Condition 2 respectively. For all subjects, except Q8 in

Condition 2, the average number of responses made in consistent transfer tests increased as a function of the number of flashes in the sample phase (left-hand column of Figs. 17 and 18). The same trend is also apparent for baseline trials in inconsistent transfer tests. However, on probe trials in inconsistent testing differences in time and rate-control performance emerged. Time-control subjects, Q5 and Q6 in Condition 1 and Q7 in Condition 2, made similar numbers of responses on all probe trials. In Condition 2, Q8 showed a systematic increase in the number of responses as flash number increased, similar to its consistent test performance in Condition 1. Conversely, rate-control subjects, Q7 and Q8 in Condition 1(Figure 17) and Q5 and Q6 in Condition 2 (Figure 18), made fewer responses on 5 and 7-flash trials than on 1 and 3-flash trials. This difference is particularly clear for Q8 in Condition 1, where the ordering of bars representing probe trial performance are reversed from that in the consistent transfer test. To quantify the trends seen in Figures 17 and 18 correlations between the number of responses made in the production phase and the number of flashes presented in the sample phase calculated. These are presented in Table 5. In consistent transfer testing, correlations between number of responses and number of flashes were positive and significant for all subjects. In inconsistent transfer testing, correlations were small and non-significant for all subjects, except Q8. Q8 showed a large and significant correlation between number of responses and number of flashes during inconsistent transfer testing, in the same direction as in consistent transfer testing (Condition 2).

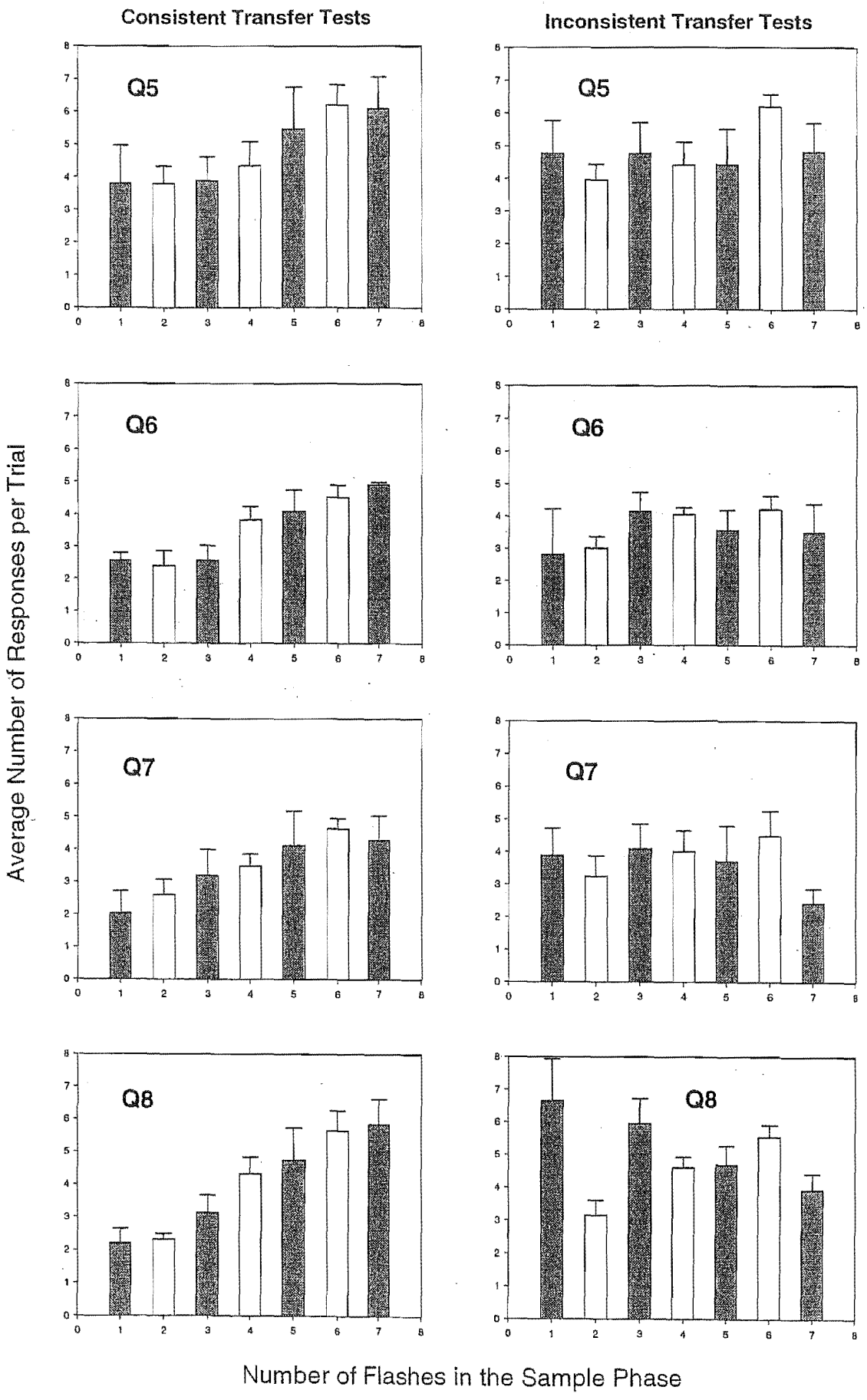

Figure 17. Number of responses (averaged over 10-sessions) and variance for each flash sequence inconsistent (left coulmn) and inconsistent (right column) transfer tests in Condition 1. White bars show baseline trials 2, 4 and 6, while grey bars show probe trials 1, 3, 5 and 7.

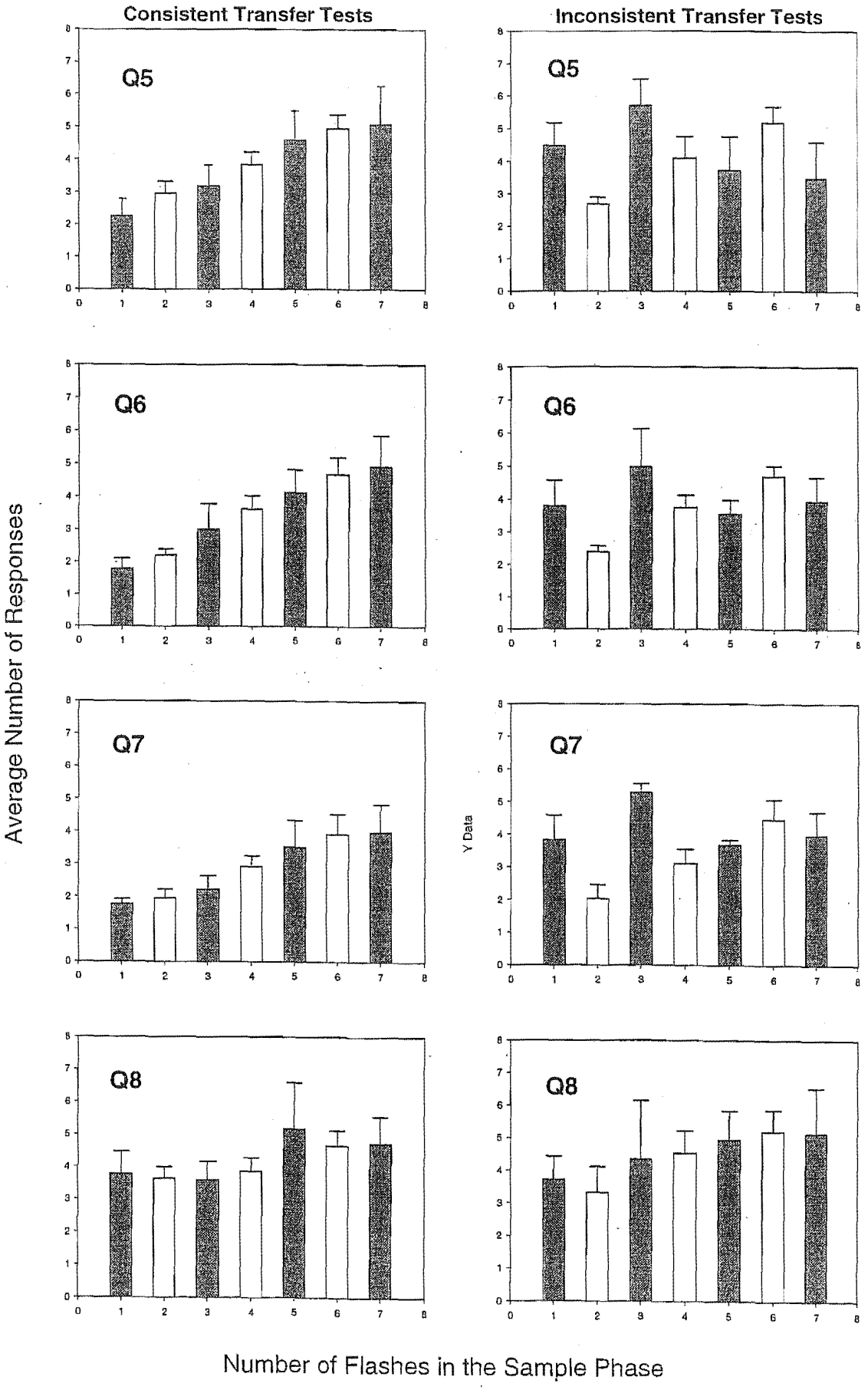

Figure 18. Number of responses (averaged over 10-sessions) and variance for each flash sequence in consistent (left column) and inconsistent (right column) transfer tests in Condition 2. White bars show baseline trials 2, 4 and 6, while grey bars show probe trials 1, 3, 5 and 7.

Table 5. Correlation between number of responses and number of flashes during consistent and inconsistent transfer tests in Condition 1 and Condition 2. * significant at $p<0.05$.

		Condition 1	Condition 2
Q5	Consistent	0.94 *	0.99 *
	Inconsistent	0.47	0.00
Q6	Consistent	0.96 *	0.99 *
	Inconsistent	0.54	0.33
Q7	Consistent	0.96 *	0.95 *
	Inconsistent	-0.26	0.28
Q8	Consistent	0.98 *	0.79 *
	Inconsistent	-0.30	0.92 *

For all subjects in consistent transfer tests in Conditions 1 and 2, response distributions for one and three-flash trials were separated, to varying extents, from response distributions for five and seven-flash trials. This suggests that subjects were discriminating between small and large numbers of flashes. To examine whether this differential performance was statistically significant, independent means t-tests were carried out. The average number of responses made on 1 and 3-flash trials for each session of testing were aggregated and compared to that for 5 and 7-flash trials. T-values and significance levels are presented in Table 6.

In consistent transfer tests all subjects made significantly more responses on 5 and 7-flash trials than on 1 and 3-flash trials. In inconsistent transfer tests the trends apparent in Figures 17 and 18 were also significant. Rate-controlled subjects made *more* responses on 1 and 3-flash trials than on 5 and 7-flash trials, showing the reverse pattern from in the consistent transfer test (note the sign changes for Q7 and Q8 in C1 and Q5 and Q6 in C2). For time-control subjects, Q5 and Q6 in Condition 1 and Q7 in Condition 2, differences between the four probe trials were not significant. Q8

(Condition 2) was the only time-control subject to show a significant difference in the same direction as in the consistent tests.

Table 6. T-values and significance levels for consistent and inconsistent transfer tests in Conditions 1 and 2. Italicized values refer to significant differences in the opposite direction to those in the consistent transfer tests.

		Condition 1		Condition 2	
		t-value	sig-level	t-value	sig-level
Q5	Consistent	-5.80	$p<0.0001$	-7.43	$p<0.0001$
	Inconsistent	0.49	ns	*4.68*	$p<0.0001$
Q6	Consistent	-9.98	$p<0.0001$	-7.64	$p<0.0001$
	Inconsistent	-0.15	ns	*2.25*	$p<0.04$
Q7	Consistent	-5.17	$p<0.0001$	-7.59	$p<0.0001$
	Inconsistent	*3.15*	$p<0.004$	0.73	ns
Q8	Consistent	-9.47	$p<0.0001$	-4.34	$p<0.001$
	Inconsistent	*6.91*	$p<0.0001$	-2.53	$p<0.01$

Comparisons between rate- and time-control performance

The results of the t-tests indicate that probe trial (but not baseline trial) performance in inconsistent transfer tests was different for rate and time-control subjects. Another area in which differences between the two procedures might emerge is in the standard deviation of response distributions. The main feature of the rate-control procedure is that sample duration was correlated with number of flashes. Given this feature one might expect subjects to base performance on temporal cues, in which case performance may reflect scalar properties of timing, such as increases in variance as the number of flashes increase (Gibbon, 1977). Figure 19 plots standard deviations as

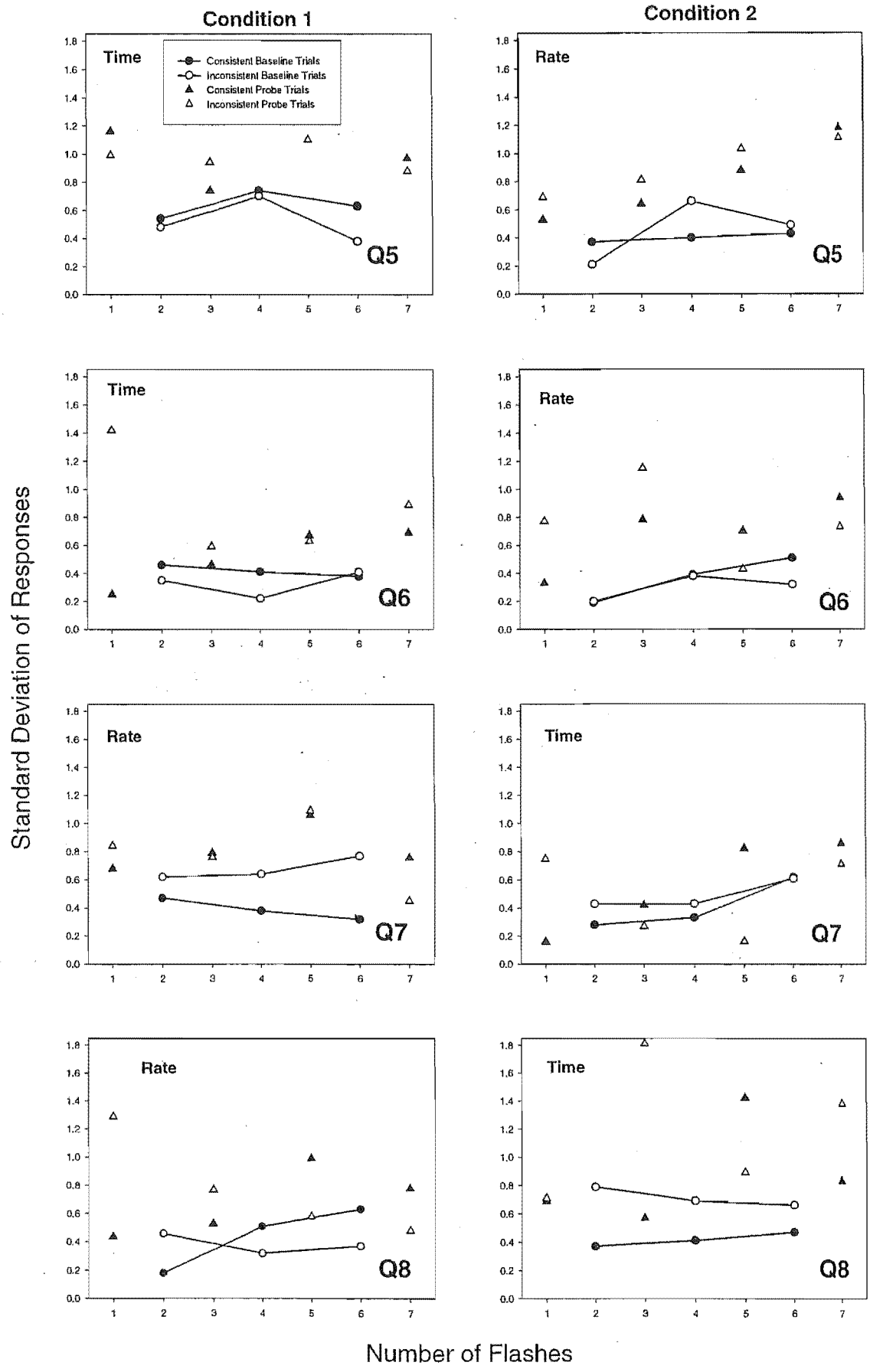

Figure 19. Standard deviation of response rates as a function of number of flashes during the sample phase for consistent (filled) and inconsistent (open) transfer tests in Conditions 1and 2. Circles represent baseline trials 2, 4 and 6 while triangles represent probe trials 1, 3, 5 and 7.

a function of number of flashes for baseline and probe trials during consistent and inconsistent testing in Conditions 1 and 2. Circles represent baseline trials while triangles represent probe trials.

In Condition 1 baseline training, standard deviations generally decreased as the number of flashes increased, while in Condition 2 they consistently increased (see Table 11 in appendix A). During testing in the first condition there were no systematic differences in standard deviations for baseline trials (circles in Figure 19) between rate and time-control procedures. One subject from each procedure showed slight increases on consistent tests and slight decreases on inconsistent tests, while the opposite was true for the remaining two subjects. In Condition 2, all subjects showed slight increases in standard deviations as the number of flashes increased on consistent tests. In inconsistent tests three subjects also had increasing functions, while Q8 had a decreasing function.

Slight differences in rate and time-control performance emerged from probe trial performance (triangles in Figure 19). On consistent transfer tests standard deviations generally increased as flash number increased for time-controlled subjects Q6 and Q7 and rate-controlled subjects Q5, Q6 and Q8. On inconsistent transfer tests there were no overall differences in standard deviations for any time-controlled subjects. However, Q8, in the rate-controlled procedure, showed a systematic decrease in standard deviations as flash number increased. Q6 and Q7 also showed slight overall decreases, however they were not systematic as for Q8. These subjects also made fewer numbers of responses as flash number increased suggesting that temporal cues may have been guiding behaviour. That Q5, who also made fewer responses on trials with larger number of flashes, showed an orderly increase in standard deviations as flash number increased is somewhat surprising.

An average response latency of 1-second was assumed when programming the presentation of flashes in the sample phase. It is possible that subjects may have produced different response latencies in the rate and time-controlled procedures, which may have affected the duration of the sample phase. To examine this average response latencies for each subject in the time and rate-controlled procedures, taken from the last five sessions of baseline training preceding consistent and inconsistent tests, were compared. There were only small differences between rate and time-controlled response latencies preceding inconsistent transfer tests. Response latency differences were much larger preceding the consistent transfer test (average of 115-ms for time-control and 62ms for rate-control). This difference was almost significant; $t=2.93$, $p<0.61$. However, the average response latency was only 15-ms greater than that predicted, which translates to only 90-ms more in the sample phase of a 6-flash trial. Thus, it is unlikely that this would have allowed subjects to use overall duration as a cue. Furthermore, if subjects had been relying on overall duration in the time-controlled procedure their performance on inconsistent transfer tests should have been much better than it was. It is also not surprising that differences in response latencies did emerge between the rate and time-controlled procedures. Smaller response latencies in the rate-controlled procedure are explained by the greater predictability of the appearance of stimuli in this procedure compared with those in the time-controlled procedure.

Summary of Results

All subjects learned to differentiate between the three baseline trials with reasonable accuracy in both conditions. Baseline performance was disrupted to

varying degrees by the introduction of consistent and inconsistent probe trials, the effects of which were reversed between Condition 1 and Condition 2. On consistent probe trials, all subjects differentiated between large and small numbers of flashes, making more responses on the former than the latter. On inconsistent probe trials, rate-control subjects made significantly more responses on trials with small numbers of flashes, in contrast to performance in consistent transfer. Time-control subjects showed no significant difference in the number of responses made on trials consisting of large or small numbers of flashes. Some subjects also appeared to differentiate between baseline and probe trials. Differences between rate and time-controlled performance also emerged in the standard deviations of responses on probe trials. Standard deviations generally increased as flash number increased for consistent transfer tests in both procedures. However, in inconsistent transfer tests three of four rate-controlled birds showed decreases in standard deviations as flash number increased, while the time-controlled subjects showed no overall trends. This may suggest that rate-controlled subjects were using temporal cues rather than numerical ones when confronted with novel probe trials.

DISCUSSION

The purpose of this study was to examine whether pigeons were able to respond in a numerical reproduction task, and if so, whether they were relying on numerical or temporal cues. This is the first time that a reproduction task such as this has been used to examine numerical ability in pigeons, and perhaps animals in general. Baseline data was orderly for all three flash sequences, with subjects showing three reasonably distinct response distributions that peaked at, or near, the reinforced number of responses (Figures 6 and 8). Subjects also showed positive transfer to novel probe trials during consistent transfer testing. All subjects made significantly more responses on 5 and 7-flash trials than on 1 and 3-flash trials. Positive correlations between number of flashes and number of responses were also obtained for all subjects. This indicates that subjects did not simply learn stereotyped response patterns, rather that there was quantitative control of behaviour by a stimulus dimension. At issue is the behavioural process responsible for this performance.

Although the time-controlled procedure was designed to promote the use of numerical cues by controlling overall duration, it did not control all the possible temporal cues that subjects could have used. In contrast, in the rate-controlled procedure overall sample duration covaried with number, and may have encouraged the use of this temporal cue. Comparisons between rate and time-controlled performance are an important source of information for understanding how subjects solved the task. In fact, differences in rate and time-controlled

performance did emerge during consistent and inconsistent transfer testing, and may suggest the use of different cues in these two procedures.

Temporal or Numerical Cues?

In this experiment it is subjects' behaviour in the production phase that is of primary interest. However, their performance in the production phase is determined by their sample phase behaviour. Presumably, subjects had to count or time the sample and then accurately reproduce that process in the production phase. By examining the sort of performance that would result from the use of various cues it is possible to evaluate which of these is most consistent with subjects' actual performance.

Perhaps the strongest indication that subjects may have been using cues other than numerical ones is the failure of subjects in the time-controlled procedure to respond accurately in the inconsistent transfer tests. Subjects' inconsistent probe trial performance in the rate-controlled procedure, compared with consistent probe trial performance, also suggests that they may have been using non-numerical cues. An increase in number of responses as flash number increased was seen for consistent probe trial performance in both procedures. However, subjects' performance in the rate-controlled procedure showed the opposite trend on inconsistent transfer tests, making a decreasing number of responses as flash number increased. This was not seen for time-controlled subjects. If subjects were attending to numerical cues then they would not make more responses on a one-flash trial than on a seven-flash trial. That all subjects did make more responses on

one and three-flash trials than on five and seven-flash trials suggests that some other type of cue may have been guiding behaviour. Past research (Roberts and Boisvert, 1998; Fetterman, 1993) has indicated that temporal cues are highly salient for many animals, including pigeons. Thus, temporal cues may have been used in preference to numerical ones.

Breukelaar and Dalrymple-Alford (1998) proposed that even when overall duration is controlled other temporal cues still co-vary with number. These include total stimulus duration, inter-event duration and the ratio of inter-event duration to total sample duration. However, none of these adequately describe the inconsistent probe trial performance seen under rate-controlled conditions. Ratios of inter-event duration to total sample duration systematically decrease as the number of flashes in the sample phase increases in both procedures (see Table 7). The addition of inconsistent probe trials would not disrupt the use of these ratios, for although the ratios differ for each flash sequence in the rate and time-controlled procedures, their ordering remains the same. Thus, the use of these ratios would only be consistent with an increase in number of responses as flash number increased. Similarly, total stimulus duration cannot account for differences in rate and time-controlled performance on inconsistent transfer tests, as it also co-varies with number of flashes in both procedures, and would only describe accurate performance.

Table 7. Programmed inter-event duration and total sample duration values for each flash sequence in the rate and time-controlled procedures. Inter-event durations are based on an average response latency of 1-second.

Time-controlled							
Number of flashes	**1**	**2**	**3**	**4**	**5**	**6**	7
Inter-event duration	9	4	2.33	1.5	1	0.66	0.43
Total sample duration	10	10	10	10	10	10	10
Rate-controlled							
Number of flashes	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Inter-event duration	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Total sample duration	2.5	5	7.5	10	12.5	15	17.5

It is interesting that a reverse ordering of responses on inconsistent tests was only seen when subjects were trained in the rate-controlled procedure and not when they were trained in the time-controlled procedure. This suggests that the cues subjects learned to rely upon during training differed between the two procedures. The important feature of the inconsistent transfer tests is that the rate of presentation of flashes changed from baseline trials to probe trials (from rate to time-controlled and vice versa). This suggests that inter-event duration may have been an important cue in inconsistent transfer test performance. In consistent transfer tests, non-numerical cues learned during training could also be used with the novel probe trials, as the manner in which flashes were presented remained the same. This is supported by the fact that all subjects, in both procedures, showed orderly increases in numbers of responses as flash number increased in consistent transfer tests. If, however, subjects did rely upon different non-numerical cues during training in the rate and time-controlled procedures, then during inconsistent

testing they may have been unable to use those same cues when presented with the probe trials. This idea is consistent with the differences seen in performance between the consistent and inconsistent transfer tests.

Whatever cue subjects used during time-controlled training resulted in them showing little discrimination between the probe trials in inconsistent transfer tests, while maintaining, to some extent, their baseline performance. Inter-event duration offers a plausible explanation for Q5's inconsistent transfer performance in the time-controlled procedure. For baseline time-controlled trials, inter-event duration decreased as flash number increased (see Table 7). For rate-controlled probe trials during inconsistent transfer testing, inter-event duration remained constant over flash sequences, and had the same value as that for time-controlled four-flash trials. If during training, time-controlled subjects relied upon inter-event duration then, during inconsistent testing, all rate-controlled probe trials would be treated as four-flash trials. This appears to be true for Q5 (see Figure 17), but not for Q7 or Q8. With respect to Q6 the results are less clear. Q6 did not show clear differentiation between the four inconsistent probe trials, with the average number of responses made on these probe trials similar to that made on four-flash baseline trials. While this does suggest that Q6 may have been basing probe trial performance on inter-event duration, matters are complicated by the fact that Q6 did not show much differentiation between the three baseline trials either. Thus, Q6's performance on all types of trials was very similar, with all response distributions peaking at, or near, four responses and little spread between them. Q8 was the only subject to make an increasing number of responses as flash number increased during inconsistent testing. Discussion of this is reserved until a later stage.

Although inter-event duration could have been used during time-controlled training, it could not be used as a cue in rate-controlled training as it remained constant over all flash sequences. If temporal cues are indeed more salient for animals they would be more likely to use overall sample duration as a cue during rate-controlled training, as it was confounded with number. During inconsistent testing however, overall sample duration could not be used as a cue, as all time-controlled flash sequences were presented within a constant programmed interval. If subjects in the rate-controlled condition had simply used overall sample duration during training, then during testing all inconsistent probe trials would have been treated as if they were rate-controlled four-flash trials. This was not the case for any subject, which indicates that some other cue must have been used.

However, it is possible that subjects' use of overall duration in baseline may have generalized to inter-event duration when presented with the inconsistent probe trials. Since overall duration did not vary for the inconsistent probe trials subjects may have used whatever temporal cue did vary with number, i.e. inter-event duration. On baseline trials sample duration increased with number, conversely, on time-controlled probe trials inter-event duration decreased as flash number increased. This explains why on inconsistent time-controlled probe trials subjects made a decreasing number of responses as flash number increased. They had learned to make an increasing number of responses in the production phase as temporal duration increased in the sample phase.

Differences in the standard deviation of responses per trial type between rate and time-controlled procedures further suggest that some subjects may have been using temporal cues. According to Scalar Expectancy Theory (Gibbon, 1977), a prominent theory of animal timing, subjects make estimates about time to

reinforcement based on a scalar timing process which is consistent with Weber's law. Scalar timing implies that the standard deviation of responses increases as a constant proportion of the mean (termed the coefficient of variation). During inconsistent transfer testing following rate-controlled training, three subjects made decreasing numbers of responses as flash number increased, and all showed decreases in standard deviations as flash number increased also. This result appears consistent with scalar expectancy theory, and if flat coefficient of variation functions were found they might imply scalar properties of timing. Figures 20 and 21 plot the coefficients of variation (CV) for baseline and probe trials respectively, during consistent and inconsistent testing. For baseline trials (Figure 20) there were no overall differences between rate and time-controlled CV functions for consistent or inconsistent transfer tests. A 2x2x3 repeated-measures ANOVA found no significant main effects or interactions between test (consistent vs inconsistent), condition (rate vs time) and number of stimuli (2, 4, 6). This indicates that CV values did not systematically change over flash sequences in both procedures.

For probe trials (Figure 21) only one time-controlled subject (Q5) showed constant CVs during inconsistent transfer, compared with three rate-controlled subjects. Similarly, in consistent tests only one time-controlled subject showed a flat

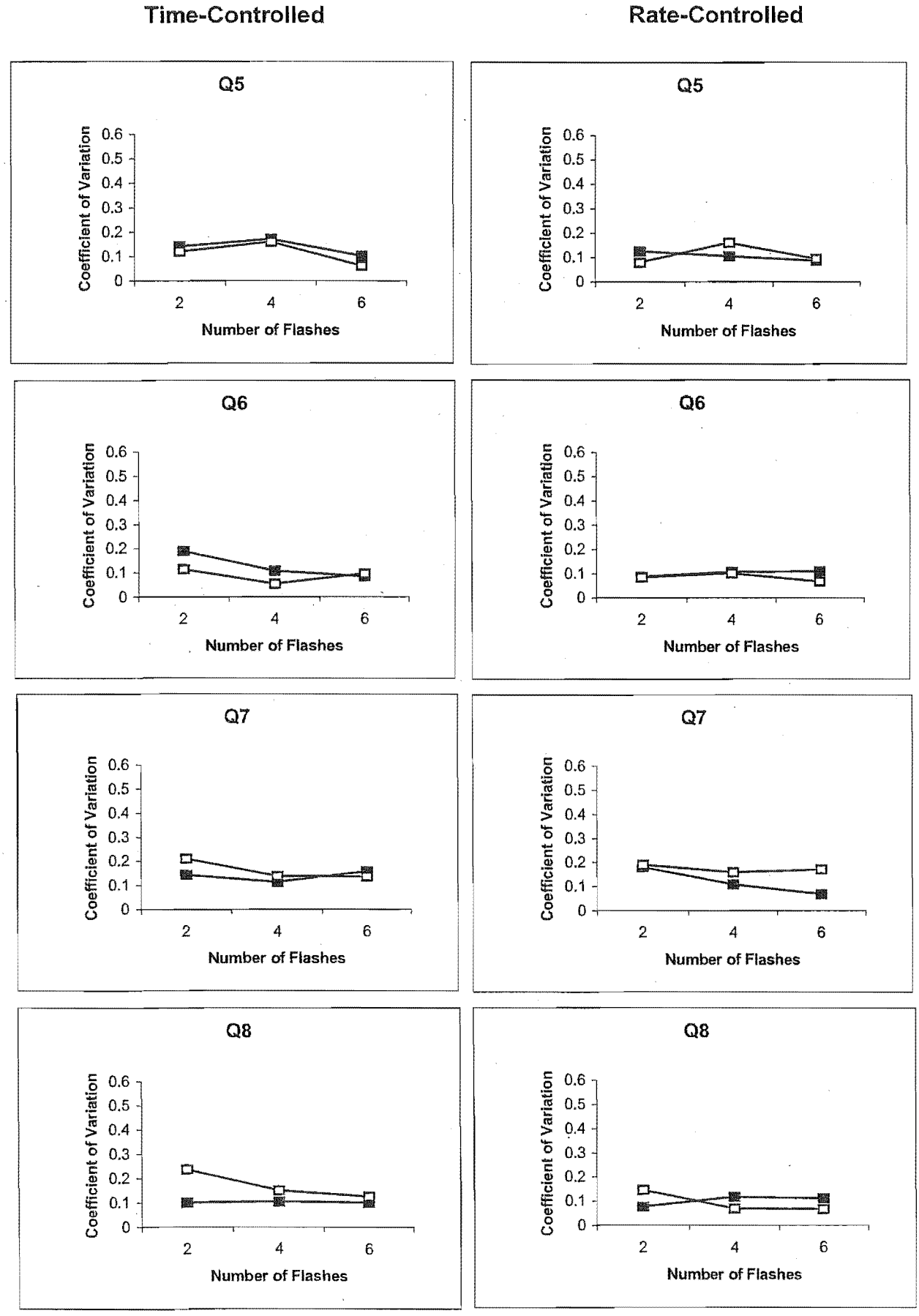

Figure 20. Coefficient of Variation plotted as a function of number of flashes for baseline trials in consistent (filled symbols) and inconsistent (open symbols) transfer tests in the rate and time-controlled procedures.

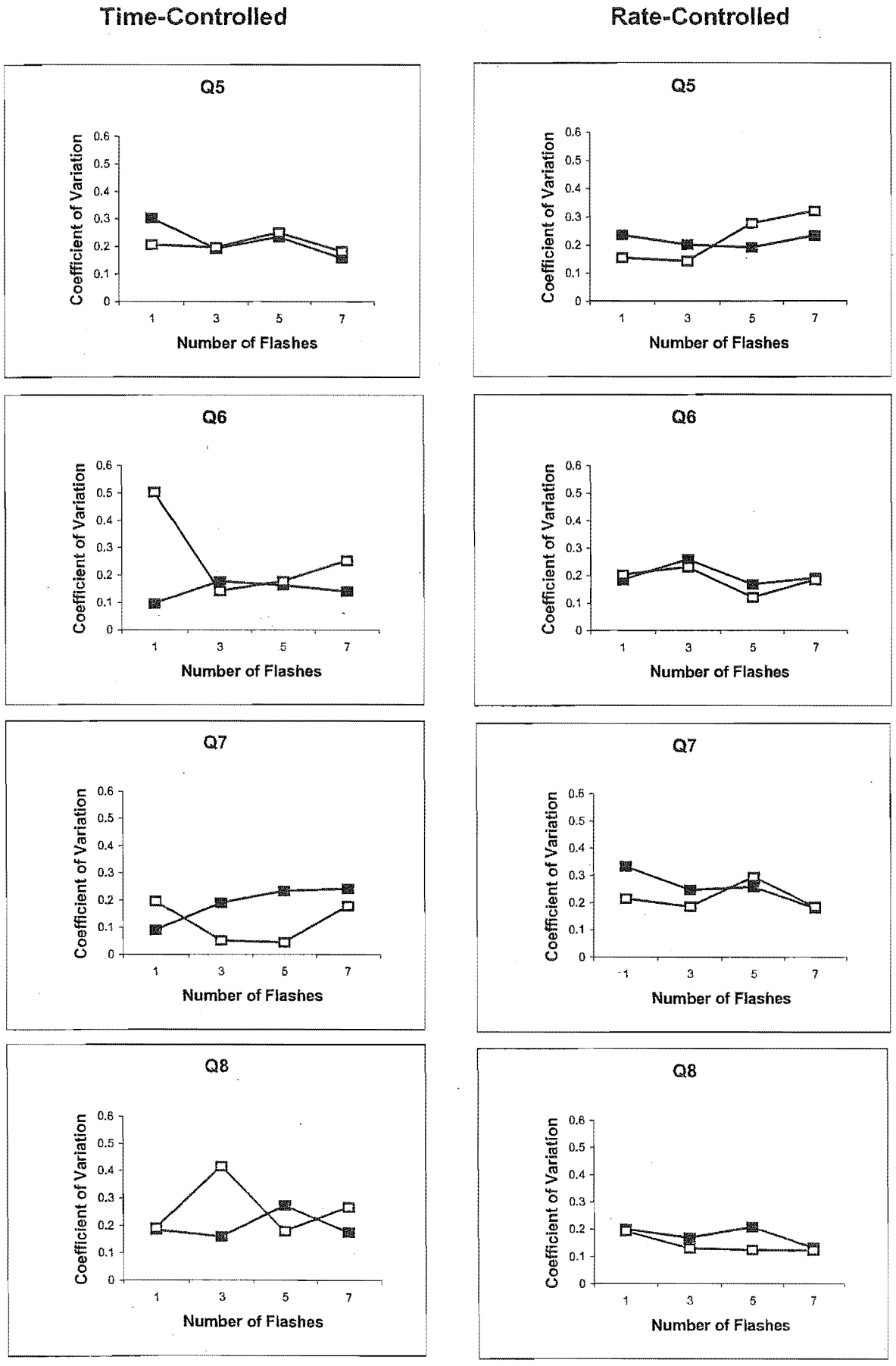

Figure 21. Coefficient of Variation plotted as a function of number of flashes for probe trials in consistent (filled symbols) and inconsistent (open symbols) transfer tests in the rate and time-controlled procedures.

CV function compared with three rate-controlled subjects. This indicates that CV functions were flatter for rate-controlled subjects than for time-controlled subjects, again suggesting that temporal cues of some form were used in the rate-controlled procedure. That Q5 appears to have a flat CV function (see Figure 21) for inconsistent transfer in the time-controlled procedure is consistent with the previous suggestion that this bird was using overall duration as a cue. For the other three birds in the time-controlled procedure, CV functions for inconsistent transfer do not appear to be flat, a finding that does not support scalar timing for these subjects. However, the results of a 2x2x4 repeated measures ANOVA conducted on probe trial CVs did not produce significant main effects or interactions either. The unsystematic variation seen in CV values means that these results are somewhat inconclusive. This variation could be the result of the relatively limited training given in each probe trial type.

Given that there are strong indications that at least some subjects were using temporal cues, the question remains as to whether there is any indication that subjects attended to numerical cues. Q8's performance in the time-controlled inconsistent transfer test tentatively suggests that this bird may have used numerical cues. It is notable that although Q8 received less baseline training than the other subjects, a programming error resulted in greater exposure to test trials (ten sessions more). This may account for the fact that Q8 was the only subject to make an increasing number of responses as flash number increased on the time-controlled inconsistent transfer test. Table 5 shows that for Q8, unlike the other subjects, the correlation between number of flashes and number of responses in the inconsistent transfer test was positive (0.92) and significant ($p<0.05$). Moreover,

THE LIBRARY
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, N.Z.

on baseline trials during inconsistent transfer CV values decreased, while for probe trials CV values varied unsystematically. This finding is not consistent with scalar expectancy theory and perhaps may indicate the use of numerical rather than temporal cues.

However, Q8's performance in the time-controlled procedure was poor compared with its rate-controlled performance in Condition 1. It may be that in Condition 2, Q8 continued to rely upon temporal cues learned during Condition 1. This would explain his poor performance on baseline and consistent probe trials. When presented with rate-controlled probe trials, in the inconsistent transfer test, the temporal cues learned in Condition 1 were accurate and Q8's performance improved. This may suggest that insufficient baseline training was given to Q8 in Condition 2, and perhaps also Q5 and Q7, both of whom showed improved performance during the inconsistent transfer test in Condition 2.

Diffusion-Confusion Model

A quantitative model is presented below, that describes subjects' performance in the response reproduction task presented in this experiment. The model involves two processes, diffusion and confusion, which when combined give predictions regarding the probability of a given number of responses being made during a particular type of trial. The model does not, however, identify the source of stimulus control. Performance predicted by the model is consistent with the use of either a temporal or numerical stimulus dimension.

After experience with a particular stimulus, organisms tend to respond to other similar stimuli in the same manner. This is termed stimulus generalisation, and can be explained by reference to a 'diffusion' process (Staddon & Reid, 1989; cf. Shepard, 1987). According to the model, reinforcement strengthens a single class of behaviour (e.g. making "two" responses in the production phase after two stimuli were presented in the sample phase), however, over time some of that strength spreads, or diffuses, to similar response classes. Initially this spread occurs only to highly similar stimuli, but as time continues since the stimulus was presented, strength spreads to more and more dissimilar stimuli. The model assumes that there is a continuum of response classes ordered by number of responses made during the production phase. Formally, the change in strength of a given response class is given by:

$$S_{j\,n+1} = (1 - 2a) \times S_{j\,n} + a(S_{j-1,\,n} + S_{j+1,\,n}) \qquad (1)$$

Where S_j is the strength of response class j, a is the rate of diffusion, and S_{j-1} and S_{j+1} are the strengths of response classes adjacent to j.

Equation 1 shows that over time some percentage (a) of the strength of a response class is lost to adjacent response classes. Assuming that the increment in response strength produced by a single reinforcer is 1.0 and that the rate of diffusion is 0.1 Equation 1 gives the following;

Table 8. Reinforcer strength at time n, n+1 and n+2 for one through eight responses on a four-flash trial.

	Reinforcer strength associated with number of centre-key responses on a four-flash trial								
Time	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
n	0	0	0	0	1	0	0	0	0
n+1	0	0	0	0.1	0.8	0.1	0	0	0
n+2	0	0	0.01	0.16	0.66	0.16	0.010	0	0

Because of the diffusion of strength to similar response classes, on some proportion of trials subjects will, for example, make three responses when presented with four-flashes. From Equation 1, three different response distributions can be produced showing the probability that a subject will make a particular number of responses given a 2, 4 or 6-flash trial.

However, the diffusion process alone is inadequate to account for performance in this task. During the production phase there is likely to be some confusion over what type of trial was presented previously. The response probability functions given by the diffusion process are likely to be changed to the extent that the three flash sequences are confused. This can be modelled as a weighted average, where the response strength distributions contribute to response probability to the extent that the three stimuli are confusable. A weighted average of the three diffusion distributions is given by:

$$P(n)_{2/4/6} = \frac{Sn_{2/4/6} + C_{2:4}\,Sn_4 + C_{2:6}\,Sn_6}{(1 + C_{2:4} + C_{2:6})} \qquad (2)$$

Where $P(n)_{2/4/6}$ is the probability of making n responses on a 2, 4 or 6-flash trial, $Sn_{2/4/6}$ is the strength of making n responses on a 2, 4 or 6-flash trial, $C_{2:4}$ is the confusion between 2 and 4-flash trials, $C_{2:6}$ is the confusion between 2 and 6-flash trials and $C_{4:6}$ is the confusion between 4 and 6-flash trials. A confusability parameter of 0 indicates no confusion between the stimuli, and increasing values denote increasing confusion.

Data for the model was averaged from the last five sessions of baseline and return to baseline in each condition. Obtained and predicted functions are shown in Figures 22-25. Values for *a* (diffusion rate), C2:4, C2:6 and C4:6, that maximised the variance accounted for (VAC) were determined using Microsoft Excel solver. These values as well as the VAC are shown in Table 9.

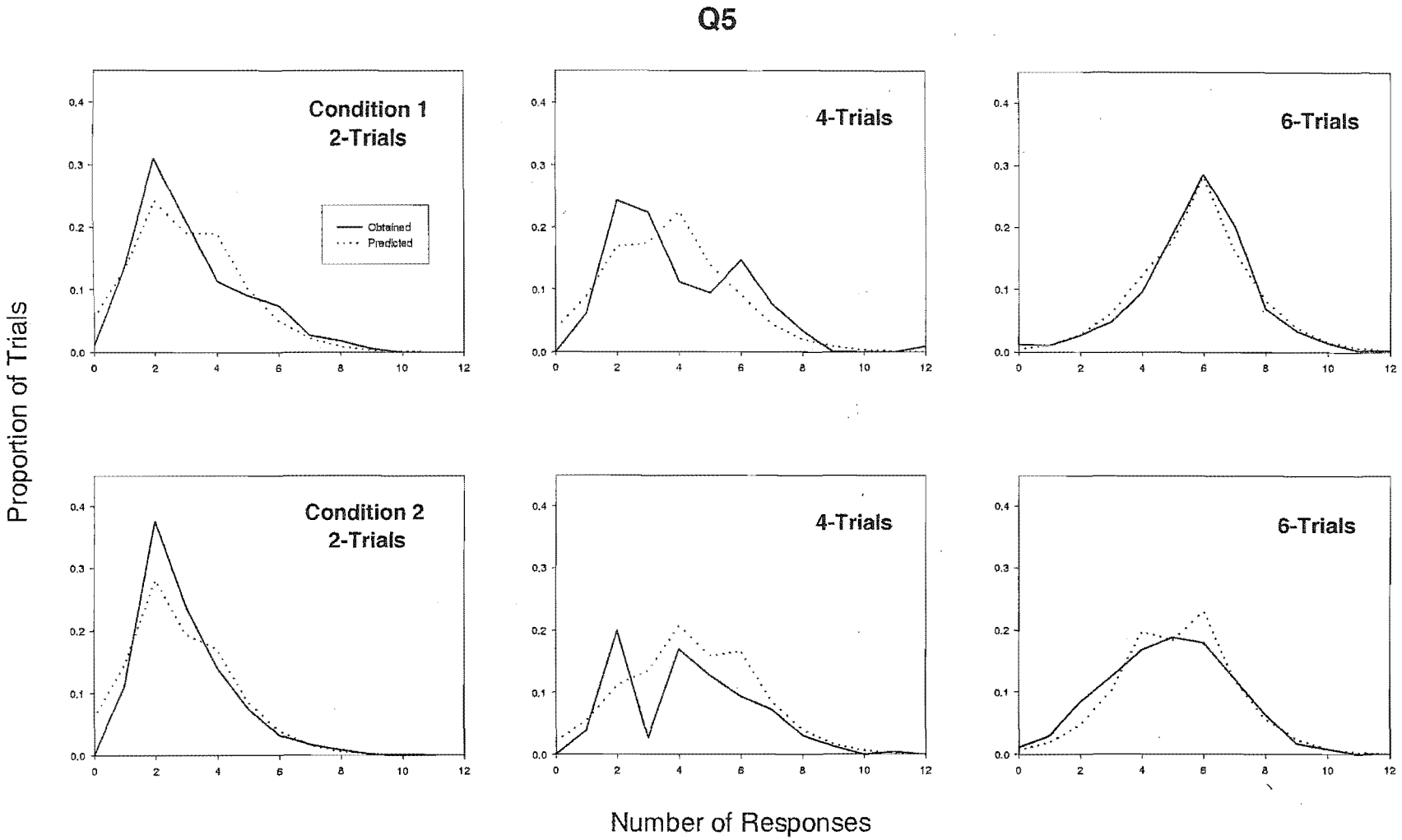

Figure 22. Obtained (solid line) and predicted (dashed line) response distributions for baseline, 2, 4 and 6-flash trials in Conditions 1 and 2 for Q5

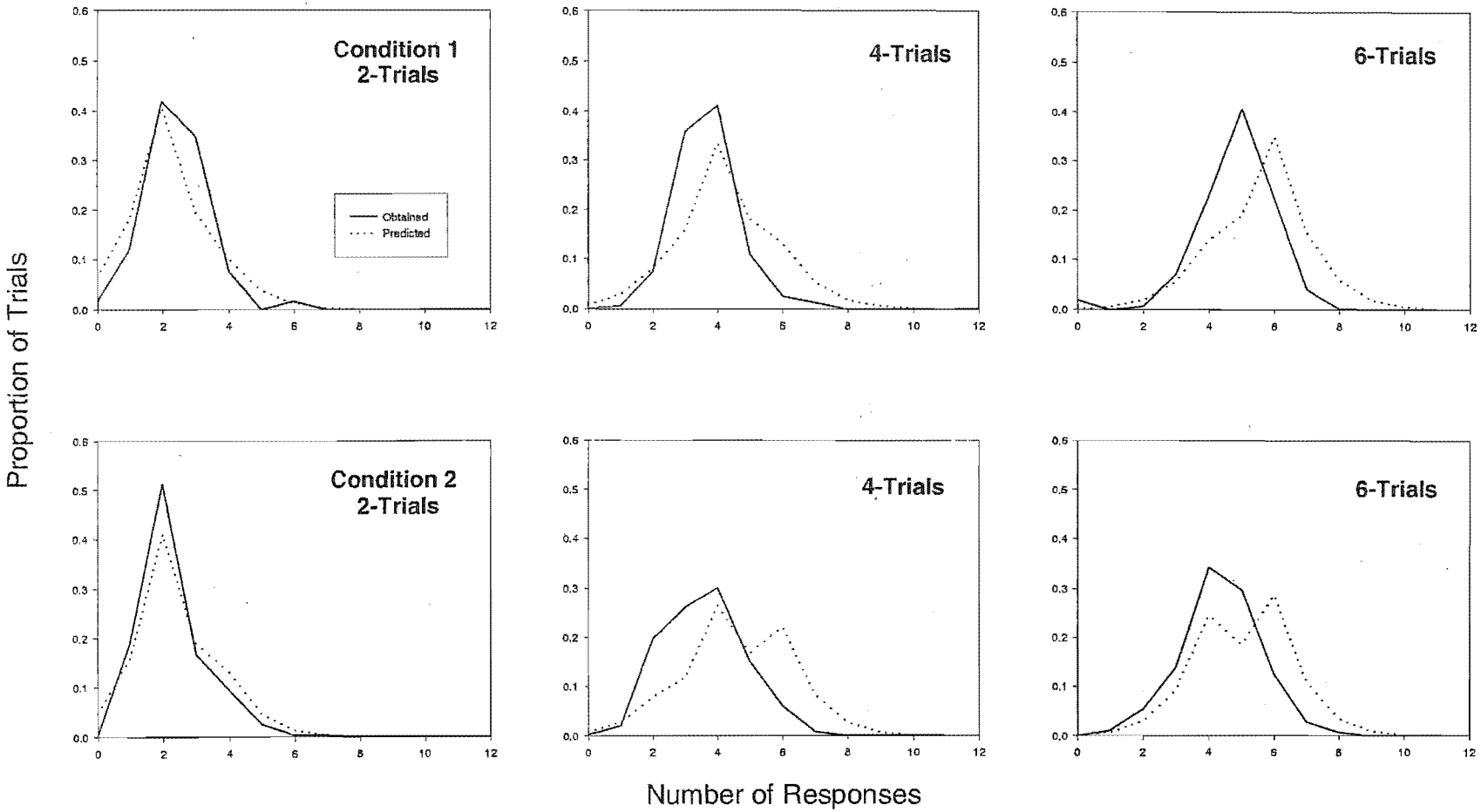

Figure 23. Obtained (solid line) and predicted (dashed line) response distributions for baseline 2, 4 and 6-flash trials in Conditions 1 and 2 for Q6.

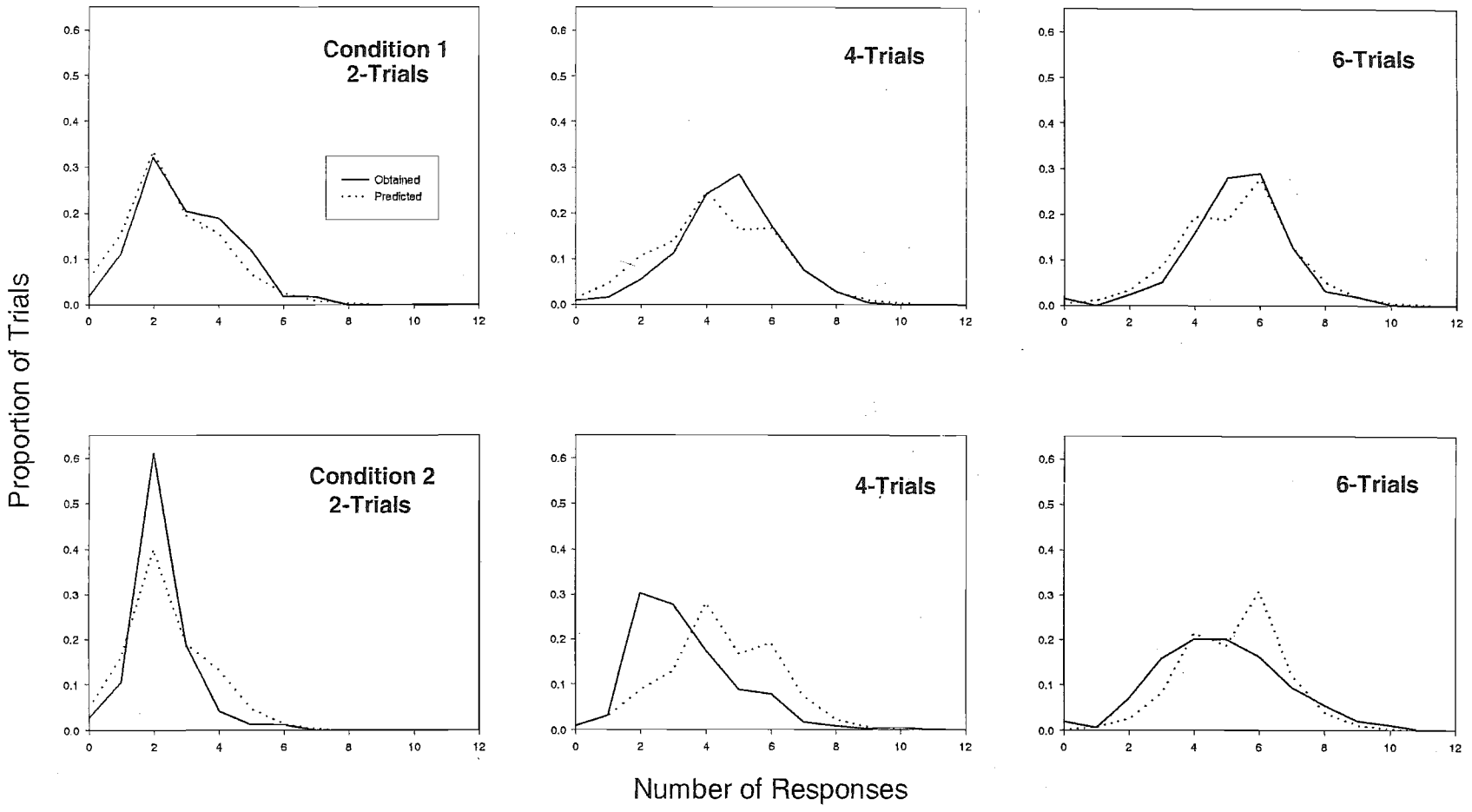

Figure 24. Obtained (solid line) and predicted (dashed line) response distributions for baseline 2, 4 and 6-flash trials in Conditions 1 and 2 for Q7.

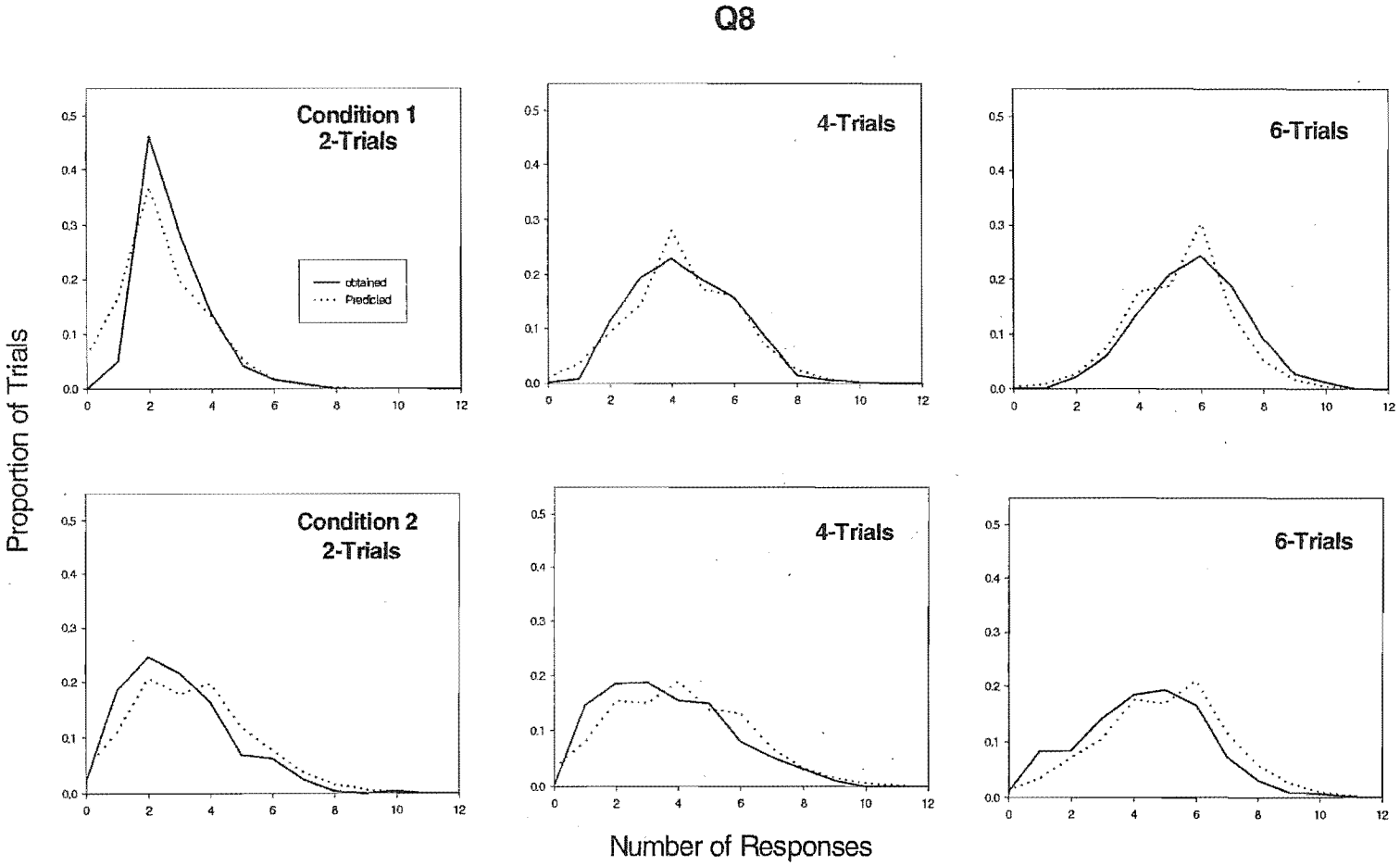

Figure 25. Obtained (solid line) and predicted (dashed line) response distributions for baseline 2, 4 and 6-flash trials in Conditions 1 and 2 for Q8.

Table 9. Best fit values for diffusion rate, confusion between 2 and 4-flash trials, 2 and 6-flash trials and 4 and 6-flash trials, and variance accounted for in Conditions 1 and 2.

		a	C2:4	C2:6	C4:6	VAC
Q5	C1	0.029	0.63	0	0.16	0.85
	C2	0.024	0.41	0	0.77	0.82
Q6	C1	0.016	0.08	0	0.25	0.73
	C2	0.012	0.21	0	0.82	0.75
Q7	C1	0.017	0.30	0	0.59	0.90
	C2	0.013	0.21	0	0.62	0.66
Q8	C1	0.016	0.19	0	0.45	0.88
	C2	0.028	0.85	0.16	0.66	0.86

Overall, the diffusion-confusion model does a good job accounting for the obtained data, with a median of 80% of the variance accounted for. Non-zero confusion parameters for two and four-flash trials and four and six-flash trials reflect the overlap in response distributions for these trials. Most subjects response distributions for two and six-flash trials showed little overlap, and this is illustrated by the confusion parameter C2:6. Only one subject had a non-zero confusion parameter for two and six-flash trials. Diffusion rates were also relatively low, ranging form 0.012 to 0.029.

Despite this there were two ways in which the model did not accurately represent the obtained data. The model always predicted a peak at the reinforced number of center key pecks, which was not always obtained (see Q6 and Q7 in Condition 2). Where confusion values were high the model also predicted two peaks

in responding: a higher peak at the reinforced number and a lower peak at the correct number of responses for the trial with which the reinforced trial was confused. For example, in Condition 2, Q6 confused four and six-flash trials to a large extent. Both predicted functions for four and six-flash trials showed primary peaks at four and six responses, and secondary peaks at six and four responses, respectively. Although the model predicted two peaks in eleven cases, the actual data almost always showed a unimodal peak. In the two cases where two peaks were obtained they were at different values than those predicted by the model. Refinements of the confusion process of the model may address the above problems and lead to a more accurate description of the data obtained in the numerical reproduction procedure.

Implications for Numerical Competence Research

Davis and Memmott (1982) proposed that subjects use numerical cues as a last resort, when no other reliable cues to reinforcement are available. Breukelaar and Dalrymple-Alford (1998) found support for this and further suggested that inaccurate temporal cues may even be used in preference to accurate numerical cues. In the present experiment, we attempted to control for temporal cues in an effort to promote numerical discrimination. However, subjects appeared to rely on temporal cues in the presence of equally valid numerical ones. This finding is consistent with that of Breukelaar and Dalrymple-Alford (1998) and Roberts and Boisvert (1998), and further indicates that subjects will use numerical cues only when forced to do so. This

illustrates that in order for performance in reproduction procedures such as this to reflect numerical discrimination, greater controls need to be taken with respect to temporal cues.

The fact that subjects appeared to be utilizing temporal rather than numerical cues makes evaluating the usefulness of a numerical reproduction paradigm for investigating animal counting difficult. These results in no way indicate that reproduction tasks cannot be used to investigate numerical ability, rather that greater care needs to be taken when deciding how the stimuli are to be presented. If sufficient controls can be taken to rule out temporal cues, and the sort of accurate performance seen in the present experiment is obtained, this type of procedure has much to say about animal numerical ability. First, numerical performance would not be due to relative numerousness judgements or subsitising (Davis & Perusse, 1988). Second, several requirements for counting would be meet. Gelman and Gallistel (1978) allow for inherent variation in the counting mechanism which animals use. Making the correct number of responses more often than not is consistent with the one to one correspondence of number of stimuli and number of responses, coupled with some degree of variability. Furthermore, accurate performance across a range of numbers would suggest that subjects represent numerosities on an ordered continuum, fulfilling the ordinality condition. Thus, accurate performance in a numerical reproduction task, when temporal cues have been ruled out, would be consistent with the formal enumerative process of counting.

The question then remains as to whether pigeons could respond accurately in the absence of temporal cues. Much of the past numerical competence research has

indicated that both pigeons and rats can utilize numerical cues when temporal ones are not available. Thus, it seems likely that various species of animals, including pigeons, could learn to differentiate numbers along a continuum, and hence to count. The onus is clearly on experimenters to develop tasks that can demonstrate that ability. The present experiment further illustrates the complexity of cues that must be considered when designing an experiment to investigate animal numerical competence.

Suggestions for Future Research

If subjects' performance in this type of production procedure is to reflect numerical discrimination clearly there needs to be greater control of temporal cues. One possibility would be to present the sample flashes at random intervals, so that inter-event duration varies within each flash sequence. Another temporal cue providing a possible confound is stimulus duration. The number of flashes in the sample phase of the present procedure is strongly correlated with the total stimulus duration.

One possible way to overcome the confound between stimulus duration and number is to change the nature of the task from response dependent to response independent. In a response independent task flashes would not be terminated when a response was made, but would be present for a predetermined interval. This would mean that the duration of each flash could be programmed so that total stimulus duration no longer varied systematically with number of flashes. However, it is not

known how this change might affect performance. One characteristic of the present experiment is that there is a one-to-one correspondence between the number of flashes and the number of responses to the stimulus in the sample phase. This 'one peck per flash' contingency may function to hone subjects' attention to the stimuli, or alternatively, each peck may produce one count (as in Gelman & Gallistel, 1978) which is used as a reference for behaviour during the production phase. It is hypothesized that subjects reproduce their own sample phase behaviour (i.e. number of pecks) during the production phase, rather than the number of flashes per se. Therefore, changing the response dependent nature of the task may result in a collapse of performance. If this did happen it would be difficult to determine whether accurate performance in a response dependent procedure was due to reliance on stimulus duration or to the one-to-one correspondence of stimuli and behaviour. On the other hand, if removing the response dependent contingency in order to adequately control for temporal cues did not affect performance, it would strongly suggest that subjects can represent numerosities on an ordered continuum.

Another feature of the present experiment is that subjects' probe trial performance generally improved over the ten sessions of testing. This raises the possibility that with a greater number of test sessions subjects performance may have been more accurate. Because of time constraints in the present study subjects received only limited exposure to test trials. A procedure in which non-reinforced probe trials are presented over an extended period of time may result in better overall performance.

Instead of having large numbers of training sessions followed by a small number of test sessions, it is possible to combine the two to create a single extended procedure. The basic idea is based on an experiment by Baer, Peterson and Sherman (1967), which demonstrated generalized imitation by children. Children were reinforced for matching particular actions made by a model. Matching of other non-reinforced actions emerged and their frequency varied with changes in reinforcement. The basic design of the present experiment would remain the same: a sample phase followed by a production phase. However, the type of trials and the number of sessions would differ. In such a procedure some trials would be reinforced while others would be non-reinforced. All reinforced and non-reinforced trials would be repeatedly presented within each session. Sessions would continue for an extended period, during which time performance on reinforced trials should gradually improve and eventually generalize to the non-reinforced trials. Probe trial performance in the present study indicated that learning occurred in the absence of differential reinforcement. This suggests that in a procedure such as the one proposed, accurate performance on non-reinforced trials would emerge over time. Table 10 shows how reinforcement might be delivered in such a procedure.

Table 10. Probability of reinforcement on trials consisting of one through eight flashes.

Number of flashes	1	2	3	4	5	6	7	8
Probability of Reinforcement	1.0	0	0	1.0	1.0	1.0	0	1.0

Trials consisting of one, four, five, six and eight flashes would always be reinforced, while trials consisting of two, three and seven flashes would never be reinforced. Relatively distinct response distributions, similar in form to those in the present experiment, should emerge over sessions for both reinforced and non-reinforced trials. As in the current experiment, the proportions of each type of trial in sessions could be altered depending on subjects' performance. Baer et. al's results suggests that pigeons would begin to differentially respond to non-reinforced trials when they are presented in the presence of other similar reinforced trials.

To take into account the need to control for temporal cues all eight flash sequences could be presented within a set interval (as in the time-controlled procedure) with varying stimulus durations and/or inter-event durations. Whether stimulus duration was varied would depend on whether the task was to be response dependent or response independent. There are merits to both procedures, and it would be interesting to compare performance under each when all other aspects of the tasks are identical. One advantage of the proposed procedure over the present one is that the design is less cumbersome and complicated. Once subjects begin training they are simply left to learn at their own rate. This avoids the problem of allowing too few sessions for performance to develop during testing, as might have been the case in the present experiment.

In summary, there is strong suggestion that subjects trained in the rate-controlled procedure based their production phase behaviour on temporal cues from the sample phase. With respect to subjects trained in the time-controlled procedure the

results also suggest the use of temporal cues. Q5's time-controlled probe trial performance is consistent with the use of inter-event duration as a cue, while Q8's inconsistent test performance also suggests that this subject was relying on some form of temporal cue. Whether Q6 and Q7 were using temporal cues in the time-controlled procedure is less clear. It appears that prior use of these cues may make it difficult for subjects to learn to use numerical ones (Q8's Condition 2 performance). If this is true, it suggests that Q7, in the time-controlled procedure, is likely to have been relying on temporal cues also. These results again illustrate the importance of ruling out temporal cues if subjects' behaviour is to reflect numerical ability.

Overall, this experiment demonstrated that pigeons were able to perform reasonably accurately on a novel numerical reproduction task. Also that performance transferred to novel stimuli when the temporal parameters of the sample phase were the same as in baseline. Pigeons' performances were well described by a diffusion-confusion model. The numerical reproduction task should prove useful in future research on numerical competence, as it provides both variety in terms of how stimuli are presented, as well as a source of behaviour

REFERENCES

Alsop, B. and Honig, W. K. (1991). Sequential stimuli and relative numerosity discriminations in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 386-395.

Baer, D. M., Peterson, R. F. and Sherman, J. A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior*, 10, 405-416.

Boysen, S. T. (1988). Commentary: kanting processes in the chimpanzee: what (and who) really counts? *Behavioral and Brain Sciences*, 11, 580.

Boysen, S. T. and Bernston, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, 103, 23-31.

Boysen, S. T. and Capaldi, E. J. (1993). *The development of numerical competence: animal and human models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Braaten, R. F. (1988). Commentary: protocounting as a last resort. *Behavioral and Brain Sciences*, 11, 581.

Brandon, P. K. (1981). A signal detection analysis of counting behavior. In M. L. Commons & J. A. Nevin (Eds.). *Quantitative analyses of behavior: Discriminative properties of reinforcement schedules* (pp 225-244). Cambridge, MA: Ballinger.

Brannon, E. M. and Terrace, H. S. (2000). Representations of the numerosities 1-9 by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, 26, 31-49.

Breukelaar, J. W. C. and Dalryample-Alford, J. C. (1998). Timing and numerical competence in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 84-97.

Bryant, P. B. and Trabasso, T. Transitive inferences and memory in young children. *Nature*, 1971, 232, 456-458.

Capaldi E. J. and Miller, D. J. (1988). Counting in rats: its functional significance and the independent processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 3-17.

Darwin, C. (1920). *The descent of man, and selection in relation to sex* (2nd ed.). New york: Appleton. (Original work published 1871)

Davis, H. (1984). Discrimination of the number three by a raccoon (*Procyon lotor*). *Animal Learning and Behavior*, 12, 409-413.

Davis, H. (1992). Transitive inferences in rats (*Rattus norvegicus*). *Journal of Comparative Psychology*, 106, 342-349.

Davis, H. (1993). Numerical competence in animals: life beyond clever Hans. In S. T. Boysen & E. J. Capaldi (Eds.). *The development of numerical competence: animal and human models* (pp 109-126). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, H. and Albert, M. (1986). Numerical discrimination by rats using sequential auditory stimuli. *Animal Learning and Behavior*, 14, 57-59.

Davis, H. and Bradford, S. A. (1991). Numerically restricted food intake in the rat in a free-feeding situation. *Animal Learning and Behavior*, 19, 215-222.

Davis, H. and Memmott, J. (1982). Counting behavior in animals: a critical evaluation. *Psychological Bulletin*, 92, 547-571.

Davis, H. and Perusse, R. (1988). Numerical competence in animals: definitional issues, current research, and a new research agenda. *Behavioral and Brain Sciences*, 11, 561-615.

Emmerton, J. (1998). Numerosity differences and effects of stimulus density on pigeons' discrimination performance. *Animal Learning and Behavior*, 26, 243-256.

Fernandes, D. M. and Church, R. M. (1982). Discrimination of the number of sequential events by rats. *Animal Learning and Behavior*, 10, 171-176.

Ferster, C. B. (1964). Arithmetic behavior in chimpanzees. *Scientific American*, 210, 98-106.

Fetterman, J. G. (1993). Numerosity discrimination: both time and number matter. *Journal of Experimental Psychology: Animal Behavior Processes*, 19, 149-164.

Gallistel, C. R. (1988). Commentary: counting versus subitizing versus the sense of number. *Behavioral and Brain Sciences*, 11, 585-586.

Gallistel, C. R. and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74.

Gelman, R. and Gallistel, C. R. *The child's understanding of number*. Cambridge, MA: Harvard University Press, 1978.

Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological Review*, 84, 279-325.

Gillan, D. J. (1981). Reasoning in the chimpanzee: II transitive inference. *Journal of Experimental Psychology*, 7, 150-164.

Hicks, L. J. (1956). An analysis of number-concept formation in the rhesus monkey. *Journal of Comparative and Physiobiological Psychology*, 49, 212-218.

Honig, W. K. and Stewart, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning and Behavior*, 17, 134-146.

Johnson, M. (1988), Commentary: out for the count. *Behavioral and Brain Sciences*, 11, 589.

Karmiloff-Smith, A. (1988). Commentary: human versus nonhuman abilities: is there a difference which really counts? *Behavioral and Brain Sciences*, 11, 589.

King, J. E. (1988). Commentary: Number concepts in animals: a multi dimensional array. *Behavioral and Brain Sciences*, 11, 590.

Luchins, A. S. and Luchins, E. H. (1988). Commentary: numbers and counting: intuitionistic and gestalt psychological viewpoints. *Behavioral and Brain Sciences*, 11, 591.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Letters to Nature*, 315, 57-59.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, 1, 109-122.

Meck, W. H. and Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320-334.

Meck, W. H., Church, R. M. and Gibbon, J. (1984). Temporal integration in duration and number discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 591-597

Miller, D, J. (1993). Do animals subsitize? In S. T. Boysen & E. J. Capaldi (Eds.). *The development of numerical competence: animal and human models* (pp 149-170). Hillsdale, NJ: Lawrence Erlbaum Associates.

Olthof, A., Iden, C. M. and Roberts, W. A. (1997). Judgements of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 325-339.

Pepperberg, I. M. (1994). Numerical competence in an African gray parrot (*Psittacus erthacus*). *Journal of Comparative psychology*, 108, 36-44.

Piaget, J. (1955). *The child's construction of reality.* London: Routledge & Kegan Paul.

Rilling, M. (1967). Number of responses as a stimulus in fixed in terval and fixed ratio schedules. *Journal of Comparative and Physiological Psychology*, 63, 60-65.

Rilling and McDiarmid (1965). Signal detection in fixed-ratio schedules. *Science*, 148, 526-527.

Roberts, W. A. (1995). Simultaneous numerical and temporal processing in the pigeon. *Current Directions in Psychological Science*, 4, 47-51.

Roberts, W. A. (1997). Does a common mechanism account for timing and counting phenomena in the pigeon? In C. M. Bradshaw &E. Szabadi (Eds.). *Time and Behaviour: Psychological and Neurobehavioural Analyses,* (pp 185-215). Amsterdam: Elsevier.

Roberts, W. A. and Boisvert, M. J. (1998). Using the peak procedure to measure timing and counting processes in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 24, 416-430.

Roberts, W. A. and Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology: Animal Behavioural Processes*, 20, 66-78.

Roberts, W. A., Macuda, T. and Broadbeck, D. R. (1995). Memory for number of light flashes in the pigeon. *Animal Learning and Behavior,* 33, 182-188.

Rumbaugh, D. M., Savage-Rumbaugh, S. and Hegel, M. T. (1987). Summation in the chimpanzee (*Pan troglodytes). Journal of Experimental Psychology: Animal Behavior Processes*, 13, 107-115.

Smirnova, A. A., Lazareva, O. F. and Zorina, Z. A. (2000). Use of number by crows: Investigation by matching and oddity learning. *Journal of the Experimental Analysis of Behavior*, 73, 163-176.

Staddon, J. E. R. and Reid, A. K. (1990). On the dynamics of generalization. *Psychological Review*, 97, 576-578.

Von Fersen, L., Wynne, C. D., Delius, J. D. and Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 334-341.

Wantanabe, S. (1998). Discrimination of 'four' and 'two' by pigeons. *The Psychological Record*, 48, 383-391.

Wilkie, D. M. Webster, J. B. and Leader L. G. (1979). Unconfounding time and number discrimination in a Mechner counting schedule. *Bulletin of the Psychonomic Society*, 13, 390-392.

Zeiler, M. D. and Hoyert, M. S. (1989). Temporal reproduction. *Journal of the Experimental Analysis of Behavior,* 52, 81-95.

Table 11. *Mean response rates and standard deviations for 2, 4 and 6-flash trials during the last 10 sessions of baseline training in Conditions 1 and 2.*

			Number of Flashes during the Production Phase		
			2	**4**	**6**
Q5	Condition 1	Mean	3.56	4.51	6.16
		SD	0.45	0.78	0.47
	Condition 2	Mean	3.26	3.89	5.60
		SD	0.29	0.42	0.58
Q6	Condition 1	Mean	2.42	3.73	4.73
		SD	0.38	0.27	0.22
	Condition 2	Mean	2.46	3.58	4.28
		SD	0.23	0.31	0.41
Q7	Condition 1	Mean	2.51	4.41	5.40
		SD	0.44	0.35	0.46
	Condition 2	Mean	2.14	2.86	4.20
		SD	0.26	0.29	0.49
Q8	Condition 1	Mean	2.74	4.09	4.97
		SD	0.41	0.29	0.30
	Condition 2	Mean	2.45	2.98	3.56
		SD	0.34	0.38	0.49

Table 12. *Mean response rates and standard deviations (10 sessions) for baseline and probe trials in consistent and inconsistent transfer tests in Condition 1.*

			Number of Flashes during the Production Phase						
			1	3	5	7	2	4	6
Q5	Consistent Transfer	Mean	3.80	3.88	5.46	6.10	3.79	4.34	6.20
		SD	1.16	0.74	1.28	0.97	0.54	0.74	0.63
	Inconsistent Transfer	Mean	4.78	4.76	4.42	4.82	3.95	4.41	6.19
		SD	0.99	0.94	1.10	0.88	0.48	0.70	0.38
Q6	Consistent Transfer	Mean	2.56	2.58	4.08	4.92	2.40	3.82	4.52
		SD	0.25	0.46	0.67	0.69	0.46	0.41	0.38
	Inconsistent Transfer	Mean	2.82	4.14	3.56	3.50	3.01	4.05	4.22
		SD	1.42	0.59	0.63	0.89	0.35	0.22	0.41
Q7	Consistent Transfer *	Mean	2.04	3.20	4.11	4.28	2.60	3.48	4.63
		SD	0.68	0.79	1.06	0.76	0.47	0.38	0.32
	Inconsistent Transfer	Mean	3.88	4.08	3.70	2.42	3.24	4.00	4.48
		SD	0.84	0.76	1.09	0.45	0.62	0.64	0.77
Q8	Consistent Transfer	Mean	2.20	3.14	4.74	5.84	2.32	4.32	5.64
		SD	0.44	0.53	0.99	0.78	0.18	0.51	0.63
	Inconsistent Transfer	Mean	6.65	5.94	4.68	3.92	3.13	4.59	5.53
		SD	1.29	0.77	0.58	0.48	0.46	0.32	0.37

* Data based on 9-session averages

Table 13. *Mean response rates and standard deviations (10 sessions) for each baseline and probe trials in consistent and inconsistent transfer tests in Condition 2.*

			Number of Flashes during the Production Phase						
			1	3	5	7	2	4	6
Q5	Consistent Transfer	Mean	2.26	3.18	4.60	5.08	2.95	3.82	4.94
		SD	0.53	0.64	0.88	1.18	0.37	0.40	0.43
	Inconsistent Transfer	Mean	4.48	5.70	3.72	3.48	2.68	4.09	5.18
		SD	0.69	0.81	1.03	1.11	0.21	0.66	0.49
Q6	Consistent Transfer	Mean	1.78	3.00	4.12	4.92	2.20	3.62	4.68
		SD	0.33	0.78	0.70	0.94	0.19	0.39	0.51
	Inconsistent Transfer	Mean	3.80	4.98	3.54	3.94	2.38	3.74	4.69
		SD	0.77	1.15	0.43	0.73	0.20	0.38	0.32
Q7	Consistent Transfer	Mean	1.76	2.22	3.52	3.54	1.94	2.92	3.91
		SD	0.16	0.42	0.82	0.86	0.28	0.33	0.62
	Inconsistent Transfer	Mean	3.83	5.27	3.67	3.97	2.02	3.11	4.45
		SD	0.75	0.27	0.16	0.71	0.43	0.43	0.61
Q8	Consistent Transfer	Mean	3.78	3.60	5.18	4.72	3.63	3.87	4.65
		SD	0.69	0.57	1.42	0.83	0.37	0.41	0.47
	Inconsistent Transfer	Mean	3.72	4.34	4.94	5.14	3.32	4.53	5.19
		SD	0.71	1.81	0.89	1.38	0.79	0.69	0.66

University of Canterbury

Private Bag 4800
Christchurch
New Zealand

Telephone: +64-3-366 7001
Facsimile: +64-3-364 2999

4 May 2000

Shasta Holland & Anthony McLean
Department of Psychology
UNIVERSITY OF CANTERBURY

Dear Shasta and Anthony

I am pleased to inform you that the Animal Ethics Committee has approved your application entitled: 2000: 15 – Counting in Pigeons: Numerical competence in a number production task.

Yours sincerely
Alan Hayward
REGISTRAR

per: Isobel Phillips
Faculty Administrator